

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:51:02 ; Search time 11.4899 Seconds
(without alignments)
1732.709 Million cell updates/sec

Title: US-09-763-902B-1
Perfect score: 2459
Sequence: 1 MRFVVALVLLNVAAGAVPL.....YIPLEKDERHQWVLLSQ 480

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2421	98.5	480	1 TGN2_HUMAN	O43493 homo sapien
2	672	27.3	357	1 TGN3_RAT	P19814 rattus norv
3	306	12.4	1087	1 NFH_MOUSE	P19246 mus musculu
4	286	11.6	831	1 NFH_RAT	P16884 rattus norv
5	272.5	11.1	1020	1 NFH_HUMAN	P12036 homo sapien
6	259	10.5	1664	1 SLPI_CLOTM	O06852 clostridium
7	257.5	10.5	1637	1 MRSP_STAAU	P80544 staphylococ
8	251.5	10.2	406	1 SR40_YEAST	P32583 saccharomyc
9	250	10.2	1253	1 DSPP_HUMAN	Q9NZW4 homo sapien
10	245	10.0	810	1 NFWM_BOVIN	O77788 bos taurus
11	241.5	9.8	704	1 NP14_RAT	P41777 rattus norv
12	236.5	9.6	915	1 NFWM_HUMAN	P07197 homo sapien
13	234	9.5	934	1 DSPP_MOUSE	P97399 mus musculu
14	232	9.4	699	1 NP14_HUMAN	Q14978 homo sapien
15	227.5	9.3	503	1 DMPL_MOUSE	O55188 mus musculu
16	226.5	9.2	488	1 CYL2_BOVIN	Q28092 bos taurus
17	226.5	9.2	1337	1 DEXT_STRDO	P39653 streptococc
18	226	9.2	489	1 DMPL_RAT	P98193 rattus norv
19	221.5	9.0	687	1 DSPP_RAT	O62598 rattus norv
20	219	8.9	1359	1 ATRX_CAEEL	Q9U7E0 caenorhabdi
21	217	8.8	700	1 TRDN_CANFA	P82179 canis famil
22	216.5	8.8	1089	1 Y553_HUMAN	Q9UKJ3 homo sapien
23	214.5	8.7	1367	1 AMYH_YEAST	P08640 saccharomyc
24	212.5	8.6	705	1 TRDN_RABIT	Q28820 oryctolagus
25	211.5	8.6	633	1 MLH_TETH	P40631 tetrahymena
26	208	8.5	1411	1 TCOF_HUMAN	Q13428 homo sapien
27	207	8.4	500	1 GAR2_SCHPO	P41891 schizosacch
28	206	8.4	400	1 RTOA_DICDI	P54681 dictyosteli
29	205.5	8.4	1210	1 AF4_HUMAN	P51825 homo sapien
30	205	8.3	644	1 NFWM_RABIT	P54938 oryctolagus
31	204	8.3	1581	1 PPRE_HUMAN	Q15648 h peroxisom
32	202.5	8.2	667	1 CYL1_BOVIN	P35662 bos taurus
33	202.5	8.2	2492	1 ATRX_HUMAN	P46100 homo sapien

34	201.5	8.2	728	1 TRDN_HUMAN	Q13061 homo sapien
35	199.5	8.1	1306	1 MSB2_YEAST	P32334 saccharomyc
36	199	8.1	559	1 ENL_HUMAN	Q03111 homo sapien
37	198.5	8.1	1770	1 PMPC_CHLTR	O84419 chlamydia t
38	196	8.0	908	1 SRCA_RABIT	P13666 oryctolagus
39	195.5	8.0	556	1 HIR3_HUMAN	Q9BW71 homo sapien
40	195.5	8.0	598	1 CYL1_HUMAN	P35663 homo sapien
41	194.5	7.9	513	1 DMF1_HUMAN	Q13316 homo sapien
42	193.5	7.9	671	1 CHS5_YEAST	Q12114 saccharomyc
43	193.5	7.9	823	1 NSP1_YEAST	P14907 saccharomyc
44	192.5	7.8	1275	1 TRP_DROME	P19334 drosophila
45	188.5	7.7	1403	1 YDF3_SCHPO	Q10475 schizosacch

ALIGNMENTS

RESULT 1
TGN2_HUMAN STANDARD; PRT: 480 AA.
ID AC O43493; O43499; O15282; Q92760; O43492; O43501;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trans-Golgi network integral membrane protein 2 precursor (Trans-
Golgi network protein TGN51) (TGN46) (TGN48) (TGN38 homolog).
GN TCOLN2 OR TGN51 OR TGN46.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND MUTAGENESIS OF TYR-430.
RC TISSUE=Liver, and placenta;
RX MEDLINE=98086273; PubMed=9422759;
RA Kain R., Angata K., Kerjaszki D., Fukuda M.;
RT "Molecular cloning and expression of a novel human trans-Golgi network
glycoprotein, TGN51, that contains multiple tyrosine-containing
motifs.";
RT J. Biol. Chem. 273:981-988(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM TGN46).
RC TISSUE=Fetal liver, and Fetal thymus;
RX MEDLINE=97063845; PubMed=8907352;
RA Ponnambalam S., Girotti M., Vaspo M.-L., Owen C.E., Perry A.C.,
RA Suganuma T., Nilsson T., Fried M., Banting G., Warren G.;
RT "Pristine homologues of rat TGN38: primary structure, expression and
functional implications.";
RL J. Cell Sci. 109:675-685(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND
FROM TRANS-GOLGI NETWORK.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-
GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL
SURFACE RETURNING VIA ENDOSOMES.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TGN46, TGN48 AND TGN51 (SHOWN
HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: ISOFORM TGN46 IS WIDELY EXPRESSED. ISOFORM
TGN51 IS MORE ABUNDANT IN FETAL LUNG AND KIDNEY. ISOFORM TGN48
IS BARELY EXPRESSED IN EMBRYONIC KIDNEY AND PROMYELOCYTIC CELLS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See: <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF027516; AAC39542.1;
EMBL; AF029316; AAB96908.1;
EMBL; AF029313; AAB96908.1; JOINED.
EMBL; AF029314; AAB96908.1; JOINED.
EMBL; AF029315; AAB96908.1; JOINED.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:51:02 ; Search time 11.4899 Seconds
(without alignments)
1732.709 Million cell updates/sec

Title: US-09-763-902B-1

Perfect score: 2459

Sequence: 1 MRFVVALVLLNVAAGAVPL.....YIPLEKDERHQWIVLLSFLQ 480

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2421	98.5	480	1 TGN2_HUMAN	O43493 homo sapien
2	672	27.3	357	1 TGN3_RAT	P19814 rattus norv
3	306	12.4	1087	1 NPH_MOUSE	P19246 mus musculu
4	286	11.6	831	1 NPH_RAT	P16884 rattus norv
5	272.5	11.1	1020	1 NPH_HUMAN	P12036 homo sapien
6	259	10.5	1664	1 SLP1_CLOTRM	P06852 clostridium
7	257.5	10.5	1637	1 MRSP_STRAU	P80544 staphylococ
8	251.5	10.2	406	1 SR40_YEAST	P32583 saccharomyc
9	250	10.2	1253	1 DSPP_HUMAN	O94244 homo sapien
10	245	10.0	810	1 NFM_BOVIN	O77788 bos taurus
11	241.5	9.8	704	1 NP14_RAT	P41777 rattus norv
12	236.5	9.6	915	1 NFM_HUMAN	P07197 homo sapien
13	234	9.5	934	1 DSPP_MOUSE	P97399 mus musculu
14	232	9.4	699	1 NP14_HUMAN	O14978 homo sapien
15	227.5	9.3	503	1 DMP1_MOUSE	O55188 mus musculu
16	226.5	9.2	488	1 CYL2_BOVIN	Q28092 bos taurus
17	226.5	9.2	1337	1 DEXT_STROD	P39653 streptococc
18	226	9.2	489	1 DMP1_RAT	P98193 rattus norv
19	221.5	9.0	687	1 DSPP_RAT	O62598 rattus norv
20	219	8.9	1359	1 ATRX_CAEEL	O94760 caenorhabdi
21	217	8.8	700	1 TRDN_CANFA	P82179 canis famli
22	216.5	8.8	1089	1 Y553_HUMAN	Q9UKJ3 homo sapien
23	214.5	8.7	1367	1 AMYH_YEAST	P08640 saccharomyc
24	212.5	8.6	705	1 TRDN_RABIT	Q28820 oryctolagus
25	211.5	8.6	633	1 MLH_TETH	P40631 tetrahymena
26	208	8.5	1411	1 TCOF_HUMAN	Q13428 homo sapien
27	207	8.4	500	1 GAR2_SCHPO	P41891 schizosacch
28	206	8.4	400	1 RTOA_DICDI	P54681 dictyostell
29	205.5	8.4	1210	1 AF4_HUMAN	P51825 homo sapien
30	205	8.3	644	1 NFM_RABIT	P54938 oryctolagus
31	204	8.3	1581	1 PPRB_HUMAN	Q15648 h peroxisom
32	202.5	8.2	667	1 CYL1_BOVIN	P35662 bos taurus
33	202.5	8.2	2492	1 ATRX_HUMAN	P46100 homo sapien

34 201.5 8.2 728 1 TRDN_HUMAN
35 199.5 8.1 1306 1 MSB2_YEAST
36 199 8.1 559 1 ENL_HUMAN
37 198.5 8.1 1770 1 PMPC_CHLTR
38 196 8.0 908 1 SRCA_RABIT
39 195.5 8.0 556 1 HIR3_HUMAN
40 195.5 8.0 598 1 CYL1_HUMAN
41 194.5 7.9 513 1 DMP1_HUMAN
42 193.5 7.9 671 1 CHS5_YEAST
43 193.5 7.9 823 1 NSPL_YEAST
44 192.5 7.8 1275 1 TRP_DROME
45 188.5 7.7 1403 1 YDF3_SCHPO

ALIGNMENTS

RESULT 1
TGN2_HUMAN
ID TGN2_HUMAN STANDARD; PRT; 480 AA.
AC O43493; O43499; O43500; O15282; Q92760; O43492; O43501;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Trans-Golgi network integral membrane protein 2 precursor (Trans-
Golgi network protein: TGN51) (TGN46) (TGN38 homolog).
GN TGN51 OR TGN51 OR TGN46.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND MUTAGENESIS OF TYR-430.
RC TISSUE=Liver, and Placenta;
RX MEDLINE-98086273; PubMed-9422759;
RT "Molecular cloning and expression of a novel human trans-Golgi network
glycoprotein, TGN51, that contains multiple tyrosine-containing
motifs";
RL J. Biol. Chem. 273:981-988(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM TGN46).
RC TISSUE=Fetal liver, and Fetal thymus;
RX MEDLINE-97063845; PubMed-8907352;
RA Ponnambalam S., Girotti M., Yaspo M.-L., Owen C.E., Perry A.C.,
Suganuma T., Nilsson T., Fried M., Banting G., Warren G.;
RT "Primate homologues of rat TGN38: primary structure, expression and
functional implications";
RL J. Cell Sci. 109:675-685(1996).
CC - FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND
FROM TRANS-GOLGI NETWORK.
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-
GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL
SURFACE RETURNING VIA ENDOSOMES.
CC - ALTERNATIVE PRODUCTS: 3 ISOFORMS: TGN46, TGN48 AND TGN51 (SHOWN
HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC - TISSUE SPECIFICITY: ISOFORM TGN46 IS WIDELY EXPRESSED. ISOFORM
TGN51 IS MORE ABUNDANT IN FETAL LUNG AND KIDNEY. ISOFORM TGN48
IS BARELY EXPRESSED IN EMBRYONIC KIDNEY AND PROMYELOCYTIC CELLS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; AF027516; AAC39542.1;
CC EMBL; AF029316; AAB96908.1;
CC EMBL; AF029313; AAB96908.1; JOINED.
CC EMBL; AF029314; AAB96908.1; JOINED.
CC EMBL; AF029315; AAB96908.1; JOINED.

```
Query Match          98.5%; Score 2421; DB 1; Length 480;
Best Local Similarity 98.5%; Pred. No. 7.7e-98;
Matches 473; Conservative. 1; Mismatches 6; Indels 0; Gaps -0;
```

Signal; Transmembrane; Glycoprotein; Repeat; Golgi stack.
 KW SIGNAL 1 17 POTENTIAL
 FT CHAIN 18 357 TRANS-GOLGI NETWORK INTEGRAL MEMBRANE
 FT PROTEIN TGN38.
 FT TRANSMEM 304 324 POTENTIAL.
 FT DOMAIN 325 357 CYTOPLASMIC TAIL, CONTAINS THE GOLGI
 RETENTION SIGNALS.
 FT DOMAIN 133 180 6 X 8 AA TANDEM REPEATS.
 FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 26 26 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MUTAGEN 348 348 S->A,D: ABOLISHES NEURABIN-I AND
 NEURABIN-II BINDING.
 FT MUTAGEN 350 350 Y->A: NO EFFECT ON NEURABIN-I AND
 NEURABIN-II BINDING.
 FT MUTAGEN 354 357 MISSING: NO EFFECT ON NEURABIN-I AND
 NEURABIN-II BINDING.
 FT SEQUENCE 357 AA; 38305 MW; 173B64C51CAD593B CRC64;
 Query Match 27.38; Score 672; DB 1; Length 357;
 Best Local Similarity 40.18; Pred. No. 6.1e-23;
 Matches 178; Conservative 48; Mismatches 122; Indels 96; Gaps 12;
 Qy 1 MREVVALLNVAAGAVPLLATESVKOEAGVRPSAGNVSTHPSLSORPGGSTRKSP-- 58
 Db 1 MOFLVALLLLSVAVARAL-----PSAKPNTVSSNNPPI 35
 Qy 59 EPQTPKSPKSSAQAQTPEDTPNKGAEAKTQKDSNKGAEAKTQKGST-S-KSGSEAQ 117
 Db 36 QPSTPLPGVDISQVKTNRPT-----DQLESDEKQDQKTVARTASVSSGVESA 86
 Qy 118 TKKD-STKSHSELQTPKDSGKSGAEAKTQKDSNKGAEAKTQKDSFGSGSEAQT 176
 Db 87 TNLNLDSSKKHPE---TADAKLETQQLLPVD-PKQKSGQKFTKDSGSPGTG----- 136
 Qy 177 DVPNKGADGQTPKDSGSGAEQDTPKDVNKGAEAKTQKDSNKGAEQGPIDGPS 236
 Db 137 DSDNTTGGD-----SNKTTGVDSKTSKGGDSNKP----- 165
 Qy 237 KSGAEQTSKDSNKPVPQPSRKDHSPISNPSNKPQLPKADTNQADKGLSPHAKT 296
 Db 166 -----TGSDN-----DKPTGGDSNKPSTKVPSTETPKIDKVLQTEKGQKPTLSKT 212
 Qy 297 ESSE-----ETDLSPQBEVKSSEPTEDVEPKAEADDTGPEGSPPKKEKMSGSASS 352
 Db 213 ESSEKLAGDSDFSLKPEKGDSSSEPTEDVETKEIEGDTPEEGSPLEBENEKVLGPS 272
 Qy 353 ENREGLTSDTSGEKDDLYPNNGSGNGSAESSHFFAYLVTAAILVAVLYTAHNNKRIIAF 412
 Db 273 ENQETLTDSMKDEKDDHYKNSGNTSRAESSHFFAYLVTAAILVAVLYTAHNNKRIIAF 332
 Qy 413 VLEGKRSKVRTRPKASDYQRLDQK 436
 Db 333 ALEGKRSKVRTRPKASDYQRLNLK 356
 RESULT 3
 ID NFH_MOUSE STANDARD; PRT; 1087 AA.
 AC P19246; Q61959;
 DT 01-NOV-1990 (rel. 16, Created)
 DT 01-AUG-1992 (rel. 23, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Neurofilament triplet H protein (200 kDa neurofilament protein)
 DE (Neurofilament heavy polypeptide) (NF-H).
 GN NFH OR NFH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=8912121513; PubMed=3220257;
 RA Julien J.-P.; Cote F.; Beaudet L.; Sidky M.; Flavell D.; Grosveld F.,
 RA Mushynski W.;
 RT "Sequence and structure of the mouse gene coding for the largest
 RT neurofilament subunit";
 RL Gene 68:307-314(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 PX MEDLINE=89089138; PubMed=3145094;
 RA Sheldman P.S.; Carden M.J.; Lees J.F.; Lazzarini R.A.;
 RT "The structure of the largest murine neurofilament protein (NF-H) as
 RT revealed by cDNA and genomic sequences";
 RL Brain Res. 464:217-231(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster; TISSUE=Brain;
 RA Carden M.J.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
 CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M24496; AAA39813.1;
 DR EMBL; M23349; AAA39813.1; JOINED.
 DR EMBL; M24494; AAA39813.1; JOINED.
 DR EMBL; M24495; AAA39813.1; JOINED.
 DR EMBL; M35131; AAA39809.1; ALT_FRAME.
 DR EMBL; Z31012; CAA83229.1;
 DR PIR; JTO368; QFM5H.
 DR PIR; A43778; A43778.
 DR MGI; 97309; Nfh.
 DR InterPro; IPR001664; IF.
 DR Pfam; PF00038; filament; 1.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 KW Repeat.
 FT DOMAIN 1 97 HEAD.
 FT DOMAIN 98 408 ROD.
 FT DOMAIN 409 1087 TAIL.
 FT DOMAIN 436 517 GLU-RICH (ACIDIC).
 FT DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
 FT DOMAIN 887 1087 GLU/LYS-RICH.
 FT DOMAIN 98 129 COIL 1A.
 FT DOMAIN 130 141 LINKER 1.
 FT DOMAIN 142 239 COIL 1B.
 FT DOMAIN 240 261 LINKER 12.
 FT DOMAIN 262 283 COIL 2A.
 FT DOMAIN 284 287 LINKER 2.
 FT DOMAIN 288 408 COIL 2B.
 FT DOMAIN 133 133 K -> QA (IN REF. 2 AND 3).
 FT DOMAIN 199 199 A -> AR (IN REF. 2 AND 3).
 FT CONFLICT 199 281 S -> T (IN REF. 2 AND 3).
 FT CONFLICT 281 281

```

FT CONFLICT 492 L -> G (IN REF. 2 AND 3).
FT CONFLICT 551 P -> PREAKSP (IN REF. 3).
FT CONFLICT 689 MISSING (IN REF. 3).
FT CONFLICT 714 G -> A (IN REF. 3).
FT CONFLICT 814 V -> M (IN REF. 2 AND 3).
FT CONFLICT 843 T -> N (IN REF. 2 AND 3).
SQ SEQUENCE 1087 AA: 57BAC76A38ED1CB9 CRC64;

Query Match 12.4%; Score 306; DB 1; Length 1087;
Best Local Similarity 26.9%; Pred. No. 8e-07;
Matches 129; Conservative 75; Mismatches 138; Indels 138; Gaps 23;

QY 13 AAGAVPPLLATESVKQEEA--GVRPSAGNVSTHPSLSORPG-----GSKSHSP 58
DB 500 AEAASPEKTKSRVKEAKSPGEAKSPGEAKS--PRAKSPGEAKSPGEAKSPGEAKSPA 558
QY 59 EPOTPKD-----SPS--KSSAEAQTPEDTPNKSQAEAKT 90
DB 559 EPKSPAEKPSPAEAKSPAEPKSPATVKSPGEAKSPSEAKSPAEPKSPAEPKSPA 616
QY 91 QKDSN---KSGAEAKTQKSTKSGSEAEATKDTKSHSELOTPKDTGKSGAEAT 146
DB 617 PAEAKSPAEPKSPAEPKSP--ATVKSPGEAKS--PSEAKSPAEPKSPAEPKSPA 670
QY 147 PEDSPNRSGAEAKTQKDSKSGSEATTKDVPNKSGADGQTPKDGSS-----KSGAEDQT 202
DB 671 PAEVKSPGEAKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 730
QY 203 PKDVPNKSGAEAKTQPKDGSN---KSGAEQGPID-----GPSKSGAEQTSKDSN 250
DB 731 PAAV--KSPGEAKSPGEAKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 788
QY 251 KVPPEPSKDKHSKTPNSDNK-----ELPKADT-----280
DB 789 KA--KSPVKEDKPPAEAKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 846
QY 281 -----NOLADKGLSPHAFKTSGET-----DLISPPQEEVKSTSEPTDVE---322
DB 847 DIRPPQVQKSPAEKAK--SPEKEAKTSEKVPKKEVSPVKEEYKAEPPKPVKEEKT 905
QY 323 ---PK-BAEDD--DTGPEEGSPPK-----BEKMKSGSASSENREGTLDSTGSEKD 368
DB 906 LPTPKTEAKESKDEAPKAPKPKVEKKEKTPTEPKDSTAEAKKEACEKKAASEEE 965

RESULT 4
NFH_RAT
ID NFH_RAT STANDARD; PRT; 831 AA.
AC P16884; Q63368;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H) (Fragment).
GN NFH OR NFH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89065087; PubMed=3143606;
RA Breen K.C., Robinson P.A., Wion D., Anderton B.H.;
RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT Identification of putative phosphorylation sites."
RL FEBS Lett. 241:213-218(1988).
RN [2]
RP SEQUENCE OF 37-831 FROM N.A.
RX MEDLINE=88309090; PubMed=2457365;
RA Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,
RA Jolles P.;
RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and

```

RT in situ detection.";
 RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
 RN [3]
 RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
 RX MEDLINE=87080760; PubMed=2878828;
 RA Robinson P.A., Wion D., Anderton B.H.;
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
 (NF-H)."
 RN FEBS Lett. 209:203-205(1986).
 RN [4]
 RP SEQUENCE OF 318-831 FROM N.A.
 RX MEDLINE=89184647; PubMed=2928342;
 RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
 RA Smolowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
 RT "Cloning of a cDNA encoding the rat high molecular weight
 neurofilament peptide (NF-H): developmental and tissue expression in
 the rat, and mapping of its human homologue to chromosomes 1 and
 22.";
 RN Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
 RL -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M37227; AAA41693.1; ALT_FRAME.
 DR EMBL; X13804; CAA32038.1; ALT_FRAME.
 DR EMBL; M21964; AAA41695.1;
 DR EMBL; J04517; AAA41692.1;
 DR PIR; A30796; A30796.
 DR PIR; A25649; A25649.
 DR PIR; B25649; B25649.
 DR PIR; S02003; S02003.
 DR InterPro; IPR001664; IF.
 DR Pfam; PF00038; filament; 1.
 DR PROSITE; PS00226; IF; 1.
 DR Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
 KW NON_TER 1
 FT DOMAIN 276 641
 FT CONFLICT 164 164
 FT CONFLICT 185 185
 FT CONFLICT 193 193
 FT CONFLICT 199 199
 FT CONFLICT 346 346
 FT CONFLICT 373 373
 FT CONFLICT 482 482
 FT CONFLICT 485 485
 FT CONFLICT 570 571
 FT CONFLICT 591 591
 FT CONFLICT 727 727
 FT CONFLICT 757 759
 FT CONFLICT 769 769
 FT CONFLICT 775 775
 SEQUENCE 831 AA: 89486 MW; 1B0973C3F13EF768 CRC64;
 51 X 3 AA TANDEM REPEATS OF K-S-P.
 L -> I (IN REF. 2).
 I -> S (IN REF. 2).
 L -> T (IN REF. 2).
 M -> T (IN REF. 2).
 K -> N (IN REF. 1).
 A -> V (IN REF. 1 AND 4).
 G -> E (IN REF. 2 AND 4).
 P -> S (IN REF. 2).
 RK -> KE (IN REF. 2 AND 4).
 P -> T (IN REF. 2 AND 4).
 A -> V (IN REF. 4).
 AAP -> GST (IN REF. 4).
 T -> L (IN REF. 2).
 R -> P (IN REF. 2 AND 4).

RESULT 6
SLP1_CLO
ID. SLP
AC. Q06

01-JUN-1994 (Rel. 29, Created)
01-JUN-1994 (Rel. 29, Last sequence update)
01-FEB-1996 (Rel. 33, Last annotation update)
Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).
GN OLPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RA MEDLINE=93209931; PubMed=8458832;
RX Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein C1pA and a protein possibly involved in attachment of the cellulosome to the cell surface."
RL J. Bacteriol. 175:1891-1899(1993).
CC -1- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC -1- SUBCELLULAR LOCATION: Cell wall.
CC -1- SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; X67506; CAA7841.1;
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH 3.
DR PROSITE; PS01072; SLH DOMAIN; 2.
KW Cell wall; S-layer; Signal; Repeat.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.
FT DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.
FT REPEAT 36 191 1.
FT REPEAT 207 363 2.
FT REPEAT 409 565 3.
FT REPEAT 607 763 4.
FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF T-P-S-D-E-P.
FT DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.
FT DOMAIN 1453 1494 SLH 1 (INCOMPLETE).
FT DOMAIN 1495 1565 SLH 2.
FT DOMAIN 1566 1625 SLH 3.
FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).
FT SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;
Query Match 10.5%; Score 259; DB 1; Length 1664;
Best Local Similarity 24.3%; Pred. No. 0.00012;
Matches 92; Conservative 62; Mismatches 178; Indels 46; Gaps 15;
QY 41 STHSLSORPGCS-TKSHPEP-QTPKSPSKSSAEATQPTDTPNKSAAEAKTQKDSNKS 98
DB 1036 SDEPTSPDEPTPEPTPEPTDTPSDEPTPS--DEPTSPDEPTSPDEPTSPDEPTSET 1093
QY 99 GAE-AKTQKSTKSGEATQTKDSTKSH----SELQTPKDTGK-----SGAEATQPE 148
DB 1094 PEPIPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTPEPTPEPTDTPSDEPTPS 1153
QY 149 DSPNRGAEAKTQKDSKSGEATQTKD--VPNKSAGDGTQPKDGSSKSG----AEQTP 203
DB 1154 DEPTSPDEPTSPDEPTSPDEPTSPDEPTPEPTDTPSDEPTSPDEPTSPDEPTSPDEPTP 1213
QY 204 KD-----VPNKSAGKQTPKSGNKSAGAEQIDGPKSGAEQTKDSNPKV 252
DB 1214 SDEPTSPDEPTPEPTDTPSDEPTSPDEPTPS--DEPTSPDEPTPS--DEPTSPDEPTPEP 1269
QY 253 VP-----EQPSKDSKSPISNPSNDKELPKADTNQLADKGLSPHAKTESGEE---TDL 304

DB 1270 IPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 1328
QY 305 ISPPQEEVKSEPTEDVEKAEADDDTGPPEGSPPKKEKMGSGASSENREGTLDSTG 364
DB 1329 -TPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 1365
QY 365 SEKDDLYPNGSGNGSAES 382
DB 1386 TPTPTTPTSGSGSGGS 1403
RESULT 7
MRSP_STA00 STANDARD; PRT; 1637 AA.
AC P80544; Q92F62;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Methicillin-resistant surface protein precursor.
GN PLS.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate 1061;
RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;
RT "pls, a large repeat-rich surface protein of methicillin resistant Staphylococcus aureus".
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;
RP 1199-1205 AND 1217-1224.
RC STRAIN=Isolate 1061;
RX MEDLINE=96270743; PubMed=8665912;
RA Hilden P., Savolainen K., Tynnelä J., Vuontola M., Kuusela P.;
RT "Purification and characterisation of a plasmin-sensitive surface protein of Staphylococcus aureus".
RL Eur. J. Biochem. 236:904-910(1996).
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; AF115379; AAD09131.1;
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR TIGRams: TIGR01167; LPXTG_anchor; 1.
DR TIGRams: TIGR01168; YSIRK_signal; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
KW Antibiotic resistance; Glycoprotein; Cell wall; Peptidoglycan-anchor; Repeat; Signal.
FT SIGNAL 1 48 POTENTIAL.
FT CHAIN 49 1601 METHICILLIN-RESISTANT SURFACE PROTEIN.
FT PROPEP 1602 1637 REMOVED BY SORTASE (POTENTIAL).
FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-[SAG].
FT SITE 1598 1602 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1601 1601 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;
Query Match 10.5%; Score 257.5; DB 1; Length 1637;
Best Local Similarity 21.6%; Pred. No. 0.00014;
Matches 85; Conservative 70; Mismatches 189; Indels 49; Gaps 8;
QY 44 PLSQRGGSTKSHP-----EPQ-----TPKSPSKSSAEQ 75


```

CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF091342; AAC36357.1;
CC InterPro; IPR001664; IF;
CC Pfam; PF00038; filament; 1.
CC PROSITE; PS00226; IF; 1.
CC Intermediate filament; Coiled coil; Neurone; Phosphorylation.
CC NON_TER 1 1
CC DOMAIN <1 296 ROD.
CC DOMAIN 297 810 TAIL.
CC DOMAIN <1 20 COIL 1A.
CC DOMAIN 21 33 LINKER 1.
CC DOMAIN 34 132 COIL 1B.
CC DOMAIN 133 149 LINKER 12.
CC DOMAIN 150 171 COIL 2A.
CC DOMAIN 172 175 LINKER 2.
CC DOMAIN 176 296 COIL 2B.
CC DOMAIN 503 582 8 X 13 AA TANDEM REPEATS.
CC SEQUENCE 810 AA; 90799 MW; B8477D85560AC3F6 CRC64;
CC
CC Query Match 10.0%; Score 245; DB 1; Length 810;
CC Best Local Similarity 28.2%; Pred. No. 0.00026;
CC Matches 111; Conservative 50; Mismatches 148; Indels 84; Gaps 21;
CC
QY 12 VAAAGAVPLLAR-----ESVKQEEAGVRPSAGNVSTHPSLSQRPGGTSKSHPEQTPK 64
DB 391 VVAAKSPVKATAPELKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE 450
QY 65 DSPSKSSAAEQTPEDTPNKGSGAEAKTKQDSSNKS-----GAEAKTKQGSTSGS-S 114
DB 451 EDEEGETERA---EGEGEAAAEAEKEEKEEKEEAEVAPKEAEAEVAPKEAEKSPVA 507
QY 115 EAQTTKSTKSHSELQTP--KDSGKSGAEATQPDSPNRSAGAEAKT-QKDSPSKSGSE 171
DB 508 KSPTTKSPAKS-PEAKSPKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSP 566
QY 172 AQT--TKDVPNKGADGQTPKDGSSKGAEADQPKD-VPNKSGAEAKTPKDGSNKSGAE 228
DB 567 AKSPAASPAKSPVBEVVKP---AEAGAEGKEQKEVEEKEEKEEKEEKEEKEEKEE 621
QY 229 QGPIGPKSGAEQTSKDPNKNVYVPEQPSRKHSPISNPSDNKELPKADTNQLADKKG 288
DB 622 EKPQVPEKKKAEKSPVKAESP--VKEEVP-----AKPV-----K 653
QY 289 LSPHAFKTESGETDLISPPQBEKSEPTEDVEP-----KEAEDDD--TGPERGSP 339
DB 654 VSP--EKEAKEE---EKPEKEKEKEVEEVEGKEEGLKESKEDIAINEVEGKEE 706
QY 340 KEE-KEKMGSGSASSENREGLSD-----STGSEK 367
DB 707 EQTEKEKGGSG---GEEEGKVVTNGLDVSPGDEK 736
CC
RESULT 11
NP14_RAT
ID NP14_RAT STANDARD; PRT; 704 AA.
AC P41777;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
DE phosphoprotein 1).

```

```

GN NOLC1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 292-309 AND 563-601.
RC TISSUE=Liver;
RX MEDLINE=92323542; PubMed=1623516;
RA Meier U.T., Biobel G.;
RT "Nopp140 shuttles on tracks between nucleolus and cytoplasm.";
RL Cell 70:127-138(1992).
RN [2]
RP INTERACTION WITH NOPS AND FIBRILLARIN.
RX MEDLINE=20143579; PubMed=10679015;
RA Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T.;
RT "Conserved composition of mammalian box H/ACA and box C/D small
RT nucleolar ribonucleoprotein particles and their interaction with the
RT common factor Nopp140.";
RL Mol. Biol. Cell 11:567-577(2000).
CC -1- FUNCTION: RELATED TO NUCLEOGENESIS, MAY PLAY A ROLE IN THE
CC MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER
CC AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC
CC GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN
CC TRANSCRIPTION CATALYZED BY RNA POLYMERASE I (BY SIMILARITY).
CC -1- SUBUNIT: Interacts with Dkl1/Nap57, Nop5/Nap65 and fibrillarin.
CC -1- SUBCELLULAR LOCATION: SHUTTLES ON CURVILINEAR TRACKS BETWEEN
CC NUCLEOLUS AND CYTOPLASM. THESE TRACKS EXTEND FROM THE DENSE
CC FIBRILLAR COMPONENT OF THE NUCLEOLUS ACROSS THE NUCLEOLAR PLASMA TO
CC A LIMITED NUMBER OF NUCLEAR PORE COMPLEXES.
CC -1- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND
CC DEPHOSPHORYLATION ON CK-II AND PKC SITES. NOPP140 IS ONE OF THE
CC MOST PHOSPHORYLATED PROTEINS IN THE CELL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M94287; AAA41718.1;
CC EMBL; M94288; AAA41719.1;
CC Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding.
CC DOMAIN 84 570 11 X 12 AA APPROXIMATE REPEATS OF AN
CC REPEAT 84 95
CC REPEAT 127 138 ACIDIC SERINE CLUSTER 1.
CC REPEAT 170 191 ACIDIC SERINE CLUSTER 2.
CC REPEAT 231 242 ACIDIC SERINE CLUSTER 3.
CC REPEAT 274 285 ACIDIC SERINE CLUSTER 4.
CC REPEAT 335 346 ACIDIC SERINE CLUSTER 5.
CC REPEAT 373 384 ACIDIC SERINE CLUSTER 6.
CC REPEAT 434 445 ACIDIC SERINE CLUSTER 7.
CC REPEAT 479 490 ACIDIC SERINE CLUSTER 8.
CC REPEAT 524 535 ACIDIC SERINE CLUSTER 9.
CC REPEAT 559 570 ACIDIC SERINE CLUSTER 10.
CC MOD_RES 567 567 PHOSPHORYLATION (BY CK2).
CC VARIANT 150 150 MISSING (IN NOPP140B).
CC SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE483EA3 CRC64;
CC
CC Query Match 9.8%; Score 241.5; DB 1; Length 704;
CC Best Local Similarity 22.5%; Pred. No. 0.00032;
CC Matches 107; Conservative 67; Mismatches 189; Indels 113; Gaps 16;
CC
QY 15 AGAVPLLATSVKQEEA-----GVRPSAGNVSTHPSLSQRPGGTSKSH 57
DB 118 AGKAAKASESSSESESESESESESESESESESESESESESESESESESESESESE 177
QY 58 PE---PQTPKDSKSSAEATQPEDT--PNKSGAEAKTKQDSSN-KSGAEAKTKQGSTK 111
DB 178 SEDEAPQTKPRAAATAAKAPTAKQTKAPAKPGPKAPKANGAGSSSSSSSSSSSS 237

```


NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster; TISSUE=Wolarr;
 RX MEDLINE=9150835; PubMed=8995371;
 RA MacDougall M., Simmons D., Luan X., Nydegger J., Feng J.Q., Gu T.T.;
 "Dentin phosphoprotein and dentin sialoprotein are cleavage products
 expressed from a single transcript coded by a gene on human chromosome
 4. Dentin phosphoprotein DNA sequence determination.";
 J. Biol. Chem. 272:835-842(1997).
 [2]
 SEQUENCE FROM N.A., AND REVISIONS TO C-TERMINUS.
 RC STRAIN=129/SVJ; TISSUE=Liver;
 RX MEDLINE=98211969; PubMed=9545272;
 RA Feng J.Q., Luan X., Wallace J., Jing D., Ohshima T., Kulkarni A.B.,
 D'Souza R.N., Kozak C.A., MacDougall M.;
 "Genomic organization, chromosomal mapping, and promoter analysis of
 the mouse dentin sialoprotein (Dsp) gene, which codes for both
 dentin sialoprotein and dentin phosphoprotein.";
 J. Biol. Chem. 273:9457-9464(1998).
 [3]
 SEQUENCE FROM N.A.
 RC STRAIN=129/SVJ; TISSUE=Liver;
 RA Sfeir C., Butler S., Lin E., George A., Veis A.;
 "From mouse to zebrafish-dentin matrix proteins genomic
 characterization.";
 Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 TISSUE SPECIFICITY.
 RX MEDLINE=98055479; PubMed=9395101;
 RA Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,
 Hottot D., Papagerakis P., Berdal A., Butler W.T.;
 "Dentin sialoprotein (DSP) transcripts: developmentally-sustained
 expression in odontoblasts and transient expression in
 pre-ameloblasts.";
 Eur. J. Oral Sci. 105:405-413(1997).
 [5]
 TISSUE SPECIFICITY.
 RX MEDLINE=21096982; PubMed=11175790;
 RA Xiao S., Yu C., Chou X., Yuan W., Wang Y., Bu L., Fu G., Qian M.,
 Yang J., Shi Y., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z.,
 Zhao G., Kong X.;
 "Dentinogenesis imperfecta 1 with or without progressive hearing loss
 is associated with distinct mutations in DSP.";
 Nat. Genet. 27:201-204(2001).
 CC - FUNCTION: DSP may be an important factor in dentinogenesis. DSP
 may bind high amount of calcium and facilitate initial
 mineralization of dentin matrix collagen as well as regulate the
 size and shape of the crystals.
 CC - SUBCELLULAR LOCATION: Secreted (By similarity).
 CC - TISSUE SPECIFICITY: Expressed in teeth, mainly in odontoblasts and
 transiently in pre-ameloblasts. Found in the inner ear.
 CC - PTM: DSP is glycosylated.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL: U67916; AAC12787.1;
 DR EMBL: AJ002141; CA005208.1; ALT_SEQ.
 DR EMBL: AF135799; A042781.1; ALT_SEQ.
 DR MGI: MGI:109172; Dssp.
 KW Signal; Extracellular matrix; Glycoprotein; Sialic acid;
 Phosphorylation.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 934 DENTIN SIALOPHOSPHOPROTEIN.
 FT CHAIN 18 451 DENTIN SIALOPROTEIN.
 FT CHAIN 452 934 DENTIN PHOSPHOPROTEIN.
 FT DOMAIN 419 934 SER/ASP-RICH.

FT SITE 479 481 CELL ATTACHMENT SITE (POTENTIAL).
 FT MOD_RES 227 227 PHOSPHORYLATION (BY CK2) (POTENTIAL).
 FT MOD_RES 254 254 PHOSPHORYLATION (BY CK1) (POTENTIAL).
 FT MOD_RES 279 279 PHOSPHORYLATION (BY CK1) (POTENTIAL).
 FT MOD_RES 293 293 PHOSPHORYLATION (BY CK1) (POTENTIAL).
 FT MOD_RES 299 299 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
 FT MOD_RES 314 314 PHOSPHORYLATION (BY CK2) (POTENTIAL).
 FT MOD_RES 336 336 PHOSPHORYLATION (BY CK2) (POTENTIAL).
 FT MOD_RES 349 349 PHOSPHORYLATION (BY CK2) (POTENTIAL).
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 934 AA; 93901 MW; A618789D8A57249A CRC64;
 Query Match 9.5%; Score 234; DB 1; Length 934;
 Best Local Similarity 20.5%; Pred. No. 0.00086;
 Matches 71; Conservative 69; Mismatches 184; Indels 22; Gaps 7;
 QY 41 STPLSLQRPGGSTKSHPEPQTPKSPKSSABAEQTPDTPNKGAEAKTQDSSNKSQA 100
 Db 585 SNSSDSDSDSGSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSS 644
 QY 101 EAKTQKGSKSGSEAEQTKDSTKSHSELQPKDSTGKSGA--EAQTPEDSPNRSQAE 158
 Db 645 SDSS 704
 QY 159 KTKDPSKSGSEAEQTKDVPNKGAGDGTGKSGAEQTPKDPNKGAEQTPK 218
 Db 705 SDSS 761
 QY 219 DGSNKGAEQEPIDGPKSGAEQTSKSPNKKVPEQSRKDHSPINSPDNKELPKA 278
 Db 762 SNSSDSS 819
 QY 279 DTNOLADKGLSPHAFKTESGETDLISPPQEVKSEPTEDVEKAEAD-DGTGPEEGS 337
 Db 820 DSS 872
 QY 338 PPKEERKMGSGASSENREGTLSDSTGSEKDDLYPNMGSGNGSAGSS 383
 Db 873 SDSS 911
 RESULT 14
 NP14_HUMAN
 ID NP14_HUMAN STANDARD; PRT; 699 AA.
 AC Q14978; Q15030;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
 nuclear phosphoprotein) (Nop140) (Nucleolar and coiled-body
 phosphoprotein 1).
 GN NOL1 OR KIAA0035.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
 RC TISSUE=Leukemia;
 RX MEDLINE=95386590; PubMed=7657714;
 RA Pai C.-Y., Chen H.-K., Sheu H.-L., Yeh N.-H.;
 "Cell-cycle-dependent alterations of a highly phosphorylated nucleolar
 protein p130 are associated with nucleogenesis.";
 J. Cell Sci. 108:1911-1920(1995).
 RN [2]
 RP SEQUENCE OF 3-699 FROM N.A. (ISOFORM BETA).
 RC TISSUE=Bone marrow;
 RX MEDLINE=96051387; PubMed=7584026;

Job time : 17.4899 secs

RL J. Bone Miner. Res. 13:422-431(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-129/SvJ; TISSUE=Spleen;
RA Feng J.O., Traianedes K., Luan X., McDougall M.;
RT "Study of murine Dmp-1 gene function and regulation.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COULD BE INVOLVED IN THE MINERALIZATION OF EXTRACELLULAR
CC MATRIX.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN
CC ODONTOBLAST, AMELOBLAST AND CEMENTOBLAST. ALSO EXPRESSED IN BONE
CC PARTICULARLY IN OSTEOBLAST.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U65020; AAB93764.1; -.
DR EMBL; AJ242625; CAB59629.1; -.
DR MGD; MGI:94910; Dmpl.
KW Extracellular matrix; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 503 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
FT DOMAIN 41 44 POLY-PRO.
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 350 352 CELL ATTACHMENT SITE (POTENTIAL).
FT CONFLICT 35 35 D -> G (IN REF. 2).
FT CONFLICT 67 67 H -> D (IN REF. 2).
FT CONFLICT 99 99 D -> E (IN REF. 2).
FT CONFLICT 116 116 G -> Y (IN REF. 2).
FT CONFLICT 137 137 A -> T (IN REF. 2).
SQ SEQUENCE 503 AA; 54000 MW; 9E8AF9F2729F113A CRC64;

Query Match 9.3%; Score 227.5; DB 1; Length 503;
Best Local Similarity 25.8%; Pred. No. 0.00095;
Matches 97; Conservative 54; Mismatches 172; Indels 53; Gaps 13;

QY 14 AAGAVPLLATSEVKQEEAGVRSAGNVSTHPSLSQRPGGSTKSHPEPQTPKDSPKSSAE 73
DB 38 AGSPPTPTNSESESESQASPEAQANDHTSHSESGELGYDRGQ----YRPAGGLSKSTGT 94
QY 74 AQTPEPTPNKGAETKDSNKGAEAKTKGTSK-----SGSEATTKDSTSKSH 127
DB 95 GADKDDDEDSGDD--TFGDEDNGLGE-EGQWGGPSKLDSDSDSADTTQSSDSTSQEN 151
QY 128 SELQTPKDSGKGAQAQPPEDSPNRSAGAEAKTKDPSKSGSEATTKDVPNKGADGQ 187
DB 152 SAQDTSPDS-----KDQDSEDDAHSRPDA-----GDSAQHSESEQRV-----GGGSEGQ 196
QY 188 TPKDSSKSGAEDQTPKDPVNPNGSAGKQTPKDGSKNGSABEQPIDGPKSGAEQTSKD 247
DB 197 S-SHGDGSEFFDEGMQSDDPPESTRDRGHARMSSAGIRSEE-----SKGDREPTSTQD 248
QY 248 SPNKVYPEQPKDHSKPISTNPNKELPADTNQLADKGLSPHAFKTESGEETDLISP 307
DB 249 SDQSQVEFSRRKSFRRSRVSEEDYR-----GELTDS-----NSRETQSDTEDTAS- 295
QY 308 POEVKSSSEPTEDVEPEAEDDDTGPEGSPKPEKMKSGSASSENRECTLSDSTGSEK 367
DB 296 -KEESRSQEDTAESQSDSPGQDPSPSESSESEAGEPESQESSSESQEGVTSERGDNP 354
QY 368 DDLYPNGSGNGSAESS 383
DB 355 D-----NTSQTDQEDS 366

Search completed: March 4, 2003, 14:59:37

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:32 ; Search time 32.8849 Seconds

(without alignments)
3007.543 Million cell updates/sec

Title: US-09-763-902B-1

Perfect score: 2459
Sequence: 1 MRFVVALVLLNVAAGAVPL.....YIPLEKDERHQTIVLLSFL 480

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phage:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_rvirus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2215	90.1	437	4 Q96QL2	Q96ql2 homo sapien
2	1811	73.6	445	6 Q95193	Q95193 macaca fasc
3	677.5	27.6	380	11 Q63575	Q63575 rattus norv
4	668	27.2	363	11 Q62314	Q62314 mus musculu
5	664	27.0	353	11 Q62313	Q62313 mus musculu
6	289	11.8	935	16 Q932C5	Q932C5 staphylococ
7	286.5	11.7	989	16 Q99V14	Q99v14 staphylococ
8	284	11.5	913	2 O86476	O86476 staphylococ
9	284	11.5	933	2 Q35653	Q35653 staphylococ
10	283	11.5	1072	11 Q35482	Q35482 rattus norv
11	276.5	11.2	1733	2 Q9K114	Q9K114 staphylococ
12	274.5	11.2	877	16 Q99R07	Q99r07 staphylococ
13	273	11.1	511	6 Q28688	Q28688 oryctolagus
14	273	11.1	606	6 Q28687	Q28687 oryctolagus
15	268	10.9	358	4 Q96HF8	Q96hf8 homo sapien
16	268	10.9	1514	5 Q8SV55	Q8sv55 drosophila

17 266.5 10.8 1026 4 Q9UJS7 Q9Ujs7 homo sapien
18 258.5 10.5 791 13 Q9DGL1 Q9Dgl1 fugu rubrip
19 258 10.5 679 5 Q95XW8 Q95xw8 caenorhabdi
20 258 10.5 2921 5 Q9N973 Q9N973 leishmania
21 251 10.2 990 13 Q91803 Q91803 xenopus lae
22 246 10.0 688 5 Q9VNX6 Q9vnx6 drosophila
23 245 10.0 970 11 Q8VBY1 Q8vdy1 rattus norv
24 242.5 9.9 1092 2 Q70022 Q70022 staphylococ
25 238.5 9.7 813 2 Q50279 Q50279 mycoplasma
26 238.5 9.7 971 5 Q9XVS4 Q9xvs4 caenorhabdi
27 238.5 9.7 1325 5 Q9BKV7 Q9bkv7 leishmania
28 238.5 9.7 2112 5 Q9VEL9 Q9vel9 drosophila
29 238 9.7 513 10 Q9LW95 Q9lww5 nicotiana t
30 236.5 9.6 1698 2 Q8LCO0 Q8lco0 staphylococ
31 234.5 9.5 734 5 Q8TIA0 Q8tia0 dictyosteli
32 234.5 9.5 881 2 Q93MH7 Q93mh7 staphylococ
33 233.5 9.5 635 10 Q4O363 Q4o363 medicago sa
34 233.5 9.5 2083 5 Q9N435 Q9n435 caenorhabdi
35 232 9.4 700 4 Q9BUV3 Q9buv3 homo sapien
36 232 9.4 722 5 Q26893 Q26893 trypanosoma
37 231.5 9.4 351 2 Q9L8W0 Q9l8w0 mycoplasma
38 230.5 9.4 401 5 Q77301 Q77301 leishmania
39 229.5 9.3 474 5 Q8TIB6 Q8tib6 dictyosteli
40 228 9.3 3507 5 Q23587 Q23587 caenorhabdi
41 228 9.3 5038 11 Q9QYX7 Q9qyx7 mus musculu
42 227.5 9.3 931 5 Q9VF92 Q9vfv92 drosophila
43 227.5 9.3 2703 5 Q9VPL9 Q9vpl9 drosophila
44 227.5 9.3 5322 5 Q9N164 Q9n164 drosophila
45 227 9.2 406 11 Q9JL35 Q9j135 mus musculu

ALIGNMENTS

RESULT 1

Q96QL2 PRELIMINARY; PRT; 437 AA.

AC Q96QL2;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Unknown (protein for MGC:14722).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008461; AAH08461.1;
SQ SEQUENCE 437 AA; 45865 MW; AA9734041CD74EFC CRC64;

Query Match 90.1%; Score 2215; DB 4; Length 437;

Best Local Similarity 99.5%; Pred. No. 3.2e-119; Mismatches 1; Indels 0; Gaps 0;
Matches 434; Conservative 1;

QY 1 MRFVVALVLLNVAAGAVPLLATSVKQEEAGVRSAGNVSTHPSLSQRPGGSTKSHPEP 60
Db 1 MRFVVALVLLNVAAGAVPLLATSVKQEEAGVRSAGNVSTHPSLSQRPGGSTKSHPEP 60
QY 61 QTPKDSKSSAEAOPTPTPNKSGAEAKTKQDSSNKSAGAEAKTKQGSTSGSGSEAOQTK 120
Db 61 QTPKDSKSSAEAOPTPTPNKSGAEAKTKQDSSNKSAGAEAKTKQGSTSGSGSEAOQTK 120
QY 121 DSTSKSHSELQTPKDSGTSGSGAEAOPTPEPSNPSGAEAKTKDSDPSKSGSEAOQTKDVNP 180
Db 121 DSTSKSHSELQTPKDSGTSGSGAEAOPTPEPSNPSGAEAKTKDSDPSKSGSEAOQTKDVNP 180
QY 181 KSGADGQTPKDGSSKSGAEADQTPKDPNKSAGAEAKTKQDPSNKSAGAEAGQPIDGPKSKGA 240
Db 181 KSGADGQTPKDGSSKSGAEADQTPKDPNKSAGAEAKTKQDPSNKSAGAEAGQPIDGPKSKGA 240

QY 241 EQTSKDSKPNKVPQPSKDKHSPISNPNKELPKADTNOLADKGLSPHAFKTESGE 300
 DB 241 EQTSKDSKPNKVPQPSKDKHSPISNPNKELPKADTNOLADKGLSPHAFKTESGE 300
 QY 301 ETDLLSPQEEVKSEPTEDVEPKAEADDDTGPEEGSPKKEKMGSGASSENREGTLS 360
 DB 301 ETDLLSPQEEVKSEPTEDVEPKAEADDDTGPEEGSPKKEKMGSGASSENREGTLS 360
 QY 361 DSTGSEKDDLPNGSGNGSASSESHFFAYLVTAAILVLYTAHNNKRIIAFVLEGRSK 420
 DB 361 DSTGSEKDDLPNGSGNGSASSESHFFAYLVTAAILVLYTAHNNKRIIAFVLEGRSK 420
 QY 421 VTRPKASDYORLDOK 436
 DB 421 VTRPKASDYORLDOK 436
 RESULT 2
 Q95193 PRELIMINARY; PRT; 445 AA.
 AC Q95193;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE Trans-golgi network integral membrane protein TGN47 precursor (TGN38 homolog)
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopitheciae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=EPIDIDYMIS;
 RX MEDLINE=97063845; PubMed=8907712;
 RA Ponnambalam S., Girotti M., Yaspo M.L., Owen C.E., Perry A.C.,
 RA Suganama T., Nilsson T., Fried M., Banting G., Warren G.,
 RT "Primate homologues of rat TGN38: primary structure, expression and functional implications";
 RL J. Cell Sci. 109:675-685(1996).
 CC - FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND FROM TRANS-GOLGI NETWORK.
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL SURFACE RETURNING VIA ENDOSOMES (BY SIMILARITY).
 DR EMBL: X94334; CAA64003.1;
 KW Signal; Transmembrane; Glycoprotein; Repeat; Golgi stack.
 FT SIGNAL 1 19
 FT CHAIN 20 445
 FT DOMAIN 20 389
 FT TRANSMEM 390 410
 FT DOMAIN 411 445
 FT DOMAIN 438 441
 FT DOMAIN 54 257
 FT REPEAT 54 67
 FT REPEAT 68 81
 FT REPEAT 82 95
 FT REPEAT 96 109
 FT REPEAT 110 123
 FT REPEAT 124 137
 FT REPEAT 138 151
 FT REPEAT 152 165
 FT REPEAT 166 179
 FT REPEAT 180 193
 FT REPEAT 194 201
 FT REPEAT 202 215
 FT REPEAT 216 229
 FT REPEAT 230 243
 FT REPEAT 244 257
 FT CARBOHYD 39 39
 FT CARBOHYD 96 96
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 445 AA; 46757 MW; C994B289A37DFAC4 CRC64;
 Query Match. 73.6%; Score 1811; DB 6; Length 445;
 Best Local Similarity 80.9%; Pred. No. 3.8e-96;
 Matches 364; Conservative 26; Mismatches 40; Indels 20; Gaps 2;
 QY 1 MRFVVALVLLNVAAGAVPLLATSVKQEEAGVYRPSAGNVSTHPSLSORPGGSTK----- 55
 DB 1 MRFVVALVLLSVAAGAAPLSDTASVKKEAGVQPSSTENISTYPTSSORPGGSTKSDPEP 60
 QY 56 -----SHPEPQTPKDSKSSAEAAQTPEDPNPKSGAEAKTKDSSNKSAGAEAKTK 106
 DB 61 HTSADSPSRSDPEPQTSKDSKSGAEAAQTPEDPNPKSGAEAKTKDSSNKSAGAEAKTK 120
 QY 107 GSTSKSGSEAAQTAKDSTKSHSELQTPKDTGKSGAEAAQTPEDPNPKSGAEAKTKDSDPS 166
 DB 121 DSSSKSGAEAKTKDSSKSGSEAAQTITKSSSKSGAEAKTKDSSKSGSETQTTKDSTS 180
 QY 167 KSGSEAAQTTKDVPNKSAGADQTPKDGSSKSGAEADQTPKDVNPKSGAEAKTKDGSNKS 226
 DB 181 KSGAEAPTTKDVPNKSAGAEADQTPKDVNPKSGAEAKTKDGSNKSAGAEAKTKDGSNKS 234
 QY 227 EEOGPDGPKSGAEAKTKDSPNPKVPEOPSRKHSKPISNPSDNKELPKADTNQLADK 286
 DB 235 EEOGPDGPKSGAEAKTKDSPNPKVPEOPSRKHSKPISNPSDNKELPKADTNQLADK 294
 QY 287 GKLSHAFKTESGEDTLSPPOEYKSEPTEDVPEKAEADDDTGPEEGSPKKEKEM 346
 DB 295 GKLSHAFKTESGEDTLSPPOEYKSEPTEDVPEKAEADDDTGPEEGSPKKEKEM 354
 QY 347 SGASSENREGTLDSTGSEKDDLPNGSGNGSASSESHFFAYLVTAAILVLYIAHNNK 406
 DB 355 SGASSENREGTLDSTGSEKDDLPNGSGNGSASSESHFFAYLVTAAILVLYIAHNNK 414
 QY 407 RKIIAFVLEGRSKVTRRPKASDYORLDOK 436
 DB 415 RKIIAFVLEGRSKVTRRPKASDYORLDOK 444
 RESULT 3
 Q63575 PRELIMINARY; PRT; 380 AA.
 ID Q63575;
 AC Q63575;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Trans golgi network (TGN) specific integral membrane protein TGN38 precursor.
 GN TGN41.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
 RX MEDLINE=92246851; PubMed=1575675;
 RA Reeves B., Wilde A., Banting G.;
 RT "Identification, molecular characterisation and immunolocalisation of an isoform of the trans-Golgi network (TGN)-specific integral membrane protein TGN38";
 RL Biochem. J. 283:313-316(1992).
 DR EMBL: X64600; CAA45884.1;
 KW Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 380
 FT POTENTIAL.
 FT TRANS GOLGI NETWORK
 FT (TGN) SPECIFIC INTEGRAL MEMBRANE PROTEIN
 FT TGN38.

in mouse.^a;
RL J. Biol. Chem. 270:14471-14476(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
CC Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND
CC FROM TRANS-GOLGI NETWORK.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-
CC GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL
CC SURFACE RETURNING VIA ENDOSOMES (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -!- MISCELLANEOUS: ALSO FOUND IN STRAINS BALB/C, C57BL/6 AND DBA/2.
DR EMBL; D50031; BAA08757.1; -
DR MGD; MGI:105080; Ttgn1.
KW Signal; Transmembrane; Glycoprotein; Repeat; Golgi stack.
FT SIGNAL 1 17 POTENTIAL:
FT CHAIN 18 353 TRANS-GOLGI NETWORK INTEGRAL MEMBRANE
FT CHAIN 18 353 PROTEIN TGN38A.
FT FT DOMAIN 18 298 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 299 319 POTENTIAL.
FT DOMAIN 320 353 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 346 349 ENDOCYTOSIS SIGNAL (BY SIMILARITY).
FT REPEAT 131 178 6 X 8 AA TANDEM REPEATS.
FT REPEAT 131 138 1.
FT REPEAT 139 146 2.
FT REPEAT 147 154 3.
FT REPEAT 155 162 4.
FT REPEAT 163 170 5.
FT REPEAT 171 178 6.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 353 AA; 95C340C2F4A21EB3 CRC64;
Query Match 27.08; Score 664; DB 11; Length 353;
Best Local Similarity 39.78; Pred. No. 9.3e-31;
Matches 177; Conservative 46; Mismatches 119; Indels 104; Gaps 10;
QY 1 MRFVVALVLLNVAAGAVPLLATESVKQEEAGVRPSAGNVTHTSLSORPGGSKSHPEP 60
DB 1 MRFQVALLLLSVAVARAL-----PSVYKRADADSGDS----- 31
QY 61 QTPKDSPSKSAEAQTPTDTPNKSGAEAKT-----QKDSNKGAEAKTKGTSKSGS 114
DB 32 QNPNQPSKQSS-----TLPSSNQVKTTRPTDGGQCKSDKKDDQKTTLLAAYSSRAES 84
QY 115 EAQTKDSTKSHSELOTPKDSKSGAEADTPDSPNRSAGAEAKTKQDSPSKSGSEAAQT 174
DB 85 GPRTAATDHLGDSRRQPEK-----TDAELNETARPLSPVNPKEK-----SDQSS 130
QY 175 TKDVPNKGADGQTPKDGSSKSGAEADTPKDVNPKSGAEKTPKDGSKNKGAEBOGPTDG 234
DB 131 TEDSGKPTGSGSGKPTGDSGK-----PTEAGSNKAT-EDDSGKS----- 169
QY 235 PSKSGAEQTSKDSKPNKVVPPQPSKQHSKPSINPSDNKELPKADTNQADKGLSPHAF 294
DB 170 -----TKVDLDKPTSKISPTSTSTKTKVQPTKEGKQPTLTS 206
QY 295 KTESGE-----ETDLISPPQEEVKSEPTEDVEPKAEADDDGPEEGSPKPEKEMSGSA 350
DB 207 KTESGETLAGSDPSFLKPEKGDKSEPTEDVETKEIEGDTPEEGSPLEENKVPGPS 266
QY 351 SSNREGTLSDSTGSEKDDLYPNGSGSAESSHFFAYLVTAALVAVLYIAHNKRKII 410
DB 267 SSNOEGTLSDSMKNEKDDLYKDSGNTSAESSHFFAYLVTAALVAVLYIAHNKRKII 326
QY 411 AFVLEGKRSKTRRPKASDYQRLDQK 436
DB 327 AFALEGKRSKTRRPKASDYQRLNLK 352
RESULT 6

Q932C5
ID Q932C5 PRELIMINARY; PRT; 935 AA.
AC Q932C5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Fibrinogen-binding protein.
GN FNB OR SAV0811.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kubara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus".
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003360; BAB56973.1; -
DR InterPro; IPR000515; BPD_transp.
DR InterPro; IPR001899; Gram_pos_anchor.
DR TIGRFAMS; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMS; TIGR01168; YSIRK_signal; 1.
DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 935 AA; 96950 MW; DC5A2D92CE3BA91C CRC64;
Query Match 11.88; Score 289; DB 16; Length 935;
Best Local Similarity 22.18; Pred. No. 6.4e-09;
Matches 79; Conservative 83; Mismatches 181; Indels 14; Gaps 5;
QY 37 AGNVSTHPSLSQRPQ--GSTKSHPE-----PQTPKDSKSAEAQTPTDTPNKSGAEAK 89
DB 535 SGGDIDKPVVPEQDEPGEIPIPEDSDSPGSDSGSDNSDSGSDSTSDSGSDSA 594
QY 90 TKQDSNKGAEAKTKGTSKSGSEAAQTTKDSTKSHSELOTPKDSKSGAEAAQTPEP 149
DB 595 SDSASASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASD 650
QY 150 SPNRSAGAEAKTKQDSPSKSGSEAAQTTPKDVNPKSGADGTPKDGSKSGAEADTPKDPNPK 209
DB 651 SDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASD 710
QY 210 SGAEKQTPKDGSKNKGAEQGPIDGPKSGAEQTSKDSKPNKVVPEQPSRKDHKPSINP 269
DB 711 SDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASD 770
QY 270 SDNKLKPAQTNQADKGLSPHAFKTESGEETDLISPPQEEVK--SEPTEDVEPKAEAD 328
DB 771 SDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASD 830
QY 329 DDTGPEEGSPKPEKMS--GSASSNREGTLSDTSGSEKDDLYPNGSGNGSAESS 383
DB 831 SDSGSDSSSDSDSTDTGSDNSDSNSDSNSDSNSDSNSDSNSDSNSDSNSDSNSDSNSD 887
RESULT 7
Q99VJ4
ID Q99VJ4 PRELIMINARY; PRT; 989 AA.
AC Q99VJ4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Fibrinogen-binding protein A, clumping factor.
GN CLFA OR SA0742.

[illegible]

```

[2]
RN      SEQUENCE FROM N.A.
RP      STRAIN-NEWMAN;
RC      Ni Eldhin D., Perkins S., Francois P., Vaudeaux P.;
RA      "Clumping factor B(Clfb), a new surface-located fibrinogen-binding
RT      adhesion of Staphylococcus aureus.";
RL      Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AJ224764; CAA12115.1;
DD      InterPro: IPR001899; Gram_pos_anchor.
DR      Pfam: PF00746; Gram_pos_anchor; 1.
DR      TIGRFAMs: TIGR01187; LPXFG_anchor; 1.
DR      TIGRFAMs: TIGR01168; YSIRK_signal; 1.
DR      PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
FT      CHAIN
FT      SEQUENCE
913 AA; 97247 MW; 38077C31F8D3E61 CRC64;

Query Match      11.5%; Score 284; DB 2; Length 913;
Best Local Similarity 21.4%; Pred. No. 1.2e-08;
Matches 78; Conservative 76; Mismatches 191; Indels 20; Gaps

QY      44  PSLQRPGGTSKSHPEPQPKDSPFSKSSAEAQTPEDTPNKSAGAEAKTKDSSNKSAGAEAK 103
Db      562  PPTPOPEPDPPEPSPDPDPDSDSDSGSDSGSDSGSDSDSDSDSDSDSDSDSDSDSDSDSD 621
QY      104  TOKGSTSKSGSEAOQTAKDSTKSKSHSELQTPKDKSTGKSGAEAOPTEDSPNKSAGAEAKTKD 163
Db      622  SEDSDSSESDDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 681
QY      164  SPKSGSEAOQTTKDVPNKSAGADQTPKDGSSKSGAEDQTPKQVPNKSAGAEAKQTPKDGSNK 223
Db      682  SDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 741
QY      224  SGAEQGPIDGPKSGAEQTSKDSKNKVVPEQPSKRDHSKPISNPDKNELPKADTNOL 283
Db      742  SDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 801
QY      284  ADGKLSPHAFKTESGEETDLISPPQEEVKSSSEPTEDVEPKAEADDDTGPEGSPPEK 343
Db      802  SD-----SDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 849
QY      344  ERWSGSASSENREGTLDSTGSEKDDLYPNGSGNGSAESHFFAYLVLTAAILVAVLYIAH 403
Db      850  KAPSNPKGEVHNKYSK---OHKTALPETGDKSENTATLFGAMW--ALLGSLLLF-- 902
QY      404  HNRKK 408
Db      903  -RRKK 906

RESULT 9
Q53653 PRELIMINARY; PRT; 933 AA.
ID      Q53653
AC      Q53653;
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DD      01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DR      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      Clumping factor.
OS      Staphylococcus aureus.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OX      Staphylococcus.
NCBI_TaxID=1280;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-NEWMAN;
RX      MEDLINE=94224142; PubMed=8170386;
RA      McDewitt D., Francois P., Vaudeaux P., Foster T.J.;
RT      "Molecular characterization of the clumping factor(fibrogen receptor
RL      of Staphylococcus aureus.";
DR      Mol. Microbiol. 11:237-248(1994).
DD      EMBL; Z18852; CAA79304.1;
DR      InterPro: IPR000515; BpD transp.

```

RESULT 9

053653

ID Q53653

AC Q53653;

DT 01-NOV-

DT 01-NOV-01 11:00

DT
DE
CLIMATE
01-JUN-70

DE C. J. Campbell
OS Stanbly

Bacteri

OC Staphyl

OX: NCBI_Ta

RN : [1]

RP.	SEQUENC
-----	---------

RC STRAIN=

RA MEDLINE
PA MoDozit

RT	¹ H Molecule	¹³ C Molecule	McDevitt
1.0	CH ₃	CH ₃	CH ₃
1.5	CH ₂	CH ₂	CH ₂
2.0	CH ₂	CH ₂	CH ₂
2.5	CH ₂	CH ₂	CH ₂
3.0	CH ₂	CH ₂	CH ₂
3.5	CH ₂	CH ₂	CH ₂
4.0	CH ₂	CH ₂	CH ₂
4.5	CH ₂	CH ₂	CH ₂
5.0	CH ₂	CH ₂	CH ₂
5.5	CH ₂	CH ₂	CH ₂
6.0	CH ₂	CH ₂	CH ₂
6.5	CH ₂	CH ₂	CH ₂
7.0	CH ₂	CH ₂	CH ₂
7.5	CH ₂	CH ₂	CH ₂
8.0	CH ₂	CH ₂	CH ₂
8.5	CH ₂	CH ₂	CH ₂
9.0	CH ₂	CH ₂	CH ₂
9.5	CH ₂	CH ₂	CH ₂
10.0	CH ₂	CH ₂	CH ₂
10.5	CH ₂	CH ₂	CH ₂
11.0	CH ₂	CH ₂	CH ₂
11.5	CH ₂	CH ₂	CH ₂
12.0	CH ₂	CH ₂	CH ₂
12.5	CH ₂	CH ₂	CH ₂
13.0	CH ₂	CH ₂	CH ₂
13.5	CH ₂	CH ₂	CH ₂
14.0	CH ₂	CH ₂	CH ₂
14.5	CH ₂	CH ₂	CH ₂
15.0	CH ₂	CH ₂	CH ₂
15.5	CH ₂	CH ₂	CH ₂
16.0	CH ₂	CH ₂	CH ₂
16.5	CH ₂	CH ₂	CH ₂
17.0	CH ₂	CH ₂	CH ₂
17.5	CH ₂	CH ₂	CH ₂
18.0	CH ₂	CH ₂	CH ₂
18.5	CH ₂	CH ₂	CH ₂
19.0	CH ₂	CH ₂	CH ₂
19.5	CH ₂	CH ₂	CH ₂
20.0	CH ₂	CH ₂	CH ₂
20.5	CH ₂	CH ₂	CH ₂
21.0	CH ₂	CH ₂	CH ₂
21.5	CH ₂	CH ₂	CH ₂
22.0	CH ₂	CH ₂	CH ₂
22.5	CH ₂	CH ₂	CH ₂
23.0	CH ₂	CH ₂	CH ₂
23.5	CH ₂	CH ₂	CH ₂
24.0	CH ₂	CH ₂	CH ₂
24.5	CH ₂	CH ₂	CH ₂
25.0	CH ₂	CH ₂	CH ₂
25.5	CH ₂	CH ₂	CH ₂
26.0	CH ₂	CH ₂	CH ₂
26.5	CH ₂	CH ₂	CH ₂
27.0	CH ₂	CH ₂	CH ₂
27.5	CH ₂	CH ₂	CH ₂
28.0	CH ₂	CH ₂	CH ₂
28.5	CH ₂	CH ₂	CH ₂
29.0	CH ₂	CH ₂	CH ₂
29.5	CH ₂	CH ₂	CH ₂
30.0	CH ₂	CH ₂	CH ₂
30.5	CH ₂	CH ₂	CH ₂
31.0	CH ₂	CH ₂	CH ₂
31.5	CH ₂	CH ₂	CH ₂
32.0	CH ₂	CH ₂	CH ₂
32.5	CH ₂	CH ₂	CH ₂
33.0	CH ₂	CH ₂	CH ₂
33.5	CH ₂	CH ₂	CH ₂
34.0	CH ₂	CH ₂	CH ₂
34.5	CH ₂	CH ₂	CH ₂
35.0	CH ₂	CH ₂	CH ₂
35.5	CH ₂	CH ₂	CH ₂
36.0	CH ₂	CH ₂	CH ₂
36.5	CH ₂	CH ₂	CH ₂
37.0			

RT of stacked

RL Mol. Mi

DR EMBL; Z

DR Interpret

10

DR InterPro: IPR001899; Gram_pos_anchor.
DR TIGRfam: TIGR01167; LPVTG_anchor; 1.
DR PROSITE: PS00402; BPD_TRANSF_INNHMER; UNKNOWN_1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 933 AA; 97058 MW; EB51AGDE2FF759F4 CRC64;

Query Match 11.5%; Score 284; DB 2; Length 933;
Best Local Similarity 22.1%; Pred. No. 1.2e-08;
Matches 80; Conservative 83; Mismatches 173; Indels 26; Gaps 5;

QY 37 AGNVTHPSLSRPG--GSKSHPE-----POTPKDPSKSSAETAQTPEDTPNKSAGAEAK 89
DB 535 SGDGIDKVPVPEOPDPGEIPIPEDSDSDPGSDSGSDNSDSDSGSDSDSDSDSDSDSD 594
QY 90 TQKDSNKSAGAEAKTQKSTSGSQAQTTKSTKSHSELQTPKDTSGKSGAEATP 149
DB 595 SDSASD 654
QY 150 SPNRSGAEAKTQKDSKSGSEATTKDVPNKGADGTPKDGSSKSGAEADOTPKDVPNK 209
DB 655 SDS 714
QY 210 SGAETPKDGNKSGAEQGPIDGPKSGAEQTSKDPNPKVVPQPKRKHRSKPSINP 269
DB 715 SDS 774
QY 270 SDNKELPKADTNQAKGLKSPHAFKTESGETDLISPPQEEVKS-----SEPTEDVE 322
DB 775 SDS 834
QY 323 PKAEADDDTGPEGSPKKEKMKSGASSENREGTLSDS-TGSEKDDLYPNKSGNGSAE 381
DB 835 SDSASD 883
QY 382 SS 383
DB 884 SN 885

RESULT 10
O35482 PRELIMINARY; PRT; 1072 AA.
AC O35482;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE High molecular-weight neurofilament.
GN NF-H.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91038277; PubMed=2230955;
RA Chin S.S., Liem R.K.;
RT "transfected rat high-molecular-weight neurofilament (NF-H)
RT coassembles with vimentin in a predominantly nonphosphorylated form.";
RL J. Neurosci. 10:3714-3726(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Chin S.S.M., Liem R.K.H.;
RA Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
RC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR EMBL: AF031879; AAB87068.1;
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; filament; 1.
DR PROSITE: PS00226; IF; 1.
KW Coiled coil; Intermediate filament.
SQ SEQUENCE 1072 AA; 115349 MW; 89A146D457A4D78B CRC64;

Query Match 11.5%; Score 283; DB 11; Length 1072;
Best Local Similarity 25.2%; Pred. No. 1.6e-08;

Matches 144; Conservative 86; Mismatches 176; Indels 166; Gaps 28;

QY 13 AAAGAVPLLATSEVKQEEAGVSPAGNVSTHPSLSORPGSKSHPEOTPKD--SP--S 68
DB 496 AEEAASPEKETSPVKEEAKSPAEAKS-----PAEAKSP-AEAKSPAEVKSPEVKSAPAE 550
QY 69 KSSAQAQTPEDTPN---KSGAEAKTQKDSNKSAGAEAKTQKSGSEAQTTKDTSTS 124
DB 551 KSPAFAKSPAEVKSATVKSAPAEAKSPAFA--KSPAFAKSP--ATVKSPEAKSPAFA 604
QY 125 KSHSELQTPKDTSGSQAQTPED--SPN--RSGAEAKTQD-----SPSKSGSA 172
DB 605 KSPAFAKSPVEA--KSPAFAKSPASVKSPEAKSPAFAKSPAEVKSATVKSPEAKSPA 662
QY 173 QTKDVPNKGADGTPKDGSS-----KSGAEADOTPK 204
DB 663 EVKSPVTVKSFAKSPAEVKSPEVKSPEAKSPAFAKSPAEVKSATVKSPEAKSPA 722
QY 205 D-----VPNKSAGAEQTPKDGSN---KSGAEQGPID-----GPSKSGAEQTSKDS 248
DB 723 EAKPPAEAKSPAFAKSPAEAKSPAFAKSPAEVKSPEAKSPAFAKSPAEVKSPEAKSPA 782
QY 249 PNKVPQSPKRDHSPKPSINPNDK-ELPKADTNQADKGL-----289
DB 783 PEKA--KSPVKEIKPPAEVKSPEAKSPAFAKSPAEVKSPEAKSPAFAKSPAEVKSPEAKSPA 840
QY 290 -----SPHAFKTESGE-----TDLISPPQEEVKSSEPTEDVE---PKAEAD 329
DB 841 PADIRSPQVKSFAKSPAEVKSPEAKSPAFAKSPAEVKSPEAKSPAFAKSPAEVKSPEAKSPA 898
QY 330 -----DTGPEGSPK-EEKMKSGASSENREGTLSDSTGSEKDDLYPNK 374
DB 899 TPATPKTEVKSPEAKSPAFAKSPAEVKSPEAKSPAFAKSPAEVKSPEAKSPAFAKSPAEVKSPEAKSPA 949
QY 375 SNGSAESSHFFAYLVAAILVAVLYIAHNNKRIIFVLEGRKSRVTRRPKASDQRLD 434
DB 950 KAAAPPEET-----PAKLGKVEAKPKKA---EDAKAKSPKSPKEPKKE 994
QY 435 QKVVLINLVPAPPKRSFLPVLTEWVPLEK 466
DB 995 E-----VPAAPKDKTKTEKTESKPKPEK 1019

RESULT 11
O9K114 PRELIMINARY; PRT; 1733 AA.
AC O9K114;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Putative cell-surface adhesin SdrF.
GN SDRF.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9491;
RA MEDLINE=20340957; PubMed=10878118;
RA McCrea K.W., Hartford O., Davis S., Ni Eidhin D., Lina G.,
RA Speziale P., Foster T.J., Hook M.;
RT "The serine-aspartate repeat (Sdr) protein family in Staphylococcus
RT epidermidis.";
RL Microbiology 146:1535-1546(2000).
DR EMBL: AF245041; AAF72509.1;
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 1733 AA; 184720 MW; D8D62EA1692FD4E8 CRC64;

Query Match 11.2%; Score 276.5; DB 2; Length 1733;
Best Local Similarity 22.0%; Pred. No. 6.3e-08;

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:58:57 ; Search time 4.18902 Seconds

(without alignments)
1761.695 Million cell updates/sec

Title: US-09-763-902B-6

Perfect score: 936

Sequence: 1 MGILLGLLLGLHFLVDYGR.....PLKATSTVKQSWDTTMDG 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	870	92.9	321	9 US-09-978-295A-52	Sequence 52, Appl
2	870	92.9	321	9 US-09-978-697-52	Sequence 52, Appl
3	870	92.9	321	9 US-09-978-192A-52	Sequence 52, Appl
4	870	92.9	321	9 US-09-953-499-2	Sequence 2, Appl
5	870	92.9	321	9 US-09-999-832A-52	Sequence 52, Appl
6	870	92.9	321	9 US-09-978-189-52	Sequence 52, Appl
7	130.5	13.9	360	9 US-09-905-291A-213	Sequence 213, App
8	130.5	13.9	360	9 US-09-902-853-213	Sequence 213, App
9	130.5	13.9	360	9 US-09-907-824-213	Sequence 213, App
10	130.5	13.9	360	9 US-09-907-841-213	Sequence 213, App
11	130.5	13.9	360	9 US-09-904-011-213	Sequence 213, App
12	130.5	13.9	360	9 US-10-174-590-28	Sequence 28, Appl
13	130.5	13.9	360	9 US-10-176-758-28	Sequence 28, Appl
14	130.5	13.9	360	9 US-10-175-737-28	Sequence 28, Appl
15	130.5	13.9	360	9 US-09-906-742-213	Sequence 213, App
16	130.5	13.9	360	9 US-10-173-706-28	Sequence 28, Appl
17	130.5	13.9	360	9 US-10-175-738-28	Sequence 28, Appl
18	130.5	13.9	360	9 US-10-175-752-28	Sequence 28, Appl
19	130.5	13.9	360	9 US-10-176-482-28	Sequence 28, Appl

20	130.5	13.9	360	9 US-10-176-757-28	Sequence 28, Appl
21	130.5	13.9	360	9 US-10-176-913-28	Sequence 28, Appl
22	130.5	13.9	360	9 US-10-180-552-28	Sequence 28, Appl
23	130.5	13.9	360	9 US-10-180-557-28	Sequence 28, Appl
24	130.5	13.9	360	9 US-09-906-838-213	Sequence 213, App
25	130.5	13.9	360	9 US-09-907-613-213	Sequence 213, App
26	130.5	13.9	360	9 US-09-907-942-213	Sequence 28, Appl
27	130.5	13.9	360	9 US-10-173-700-28	Sequence 28, Appl
28	130.5	13.9	360	9 US-10-174-572-28	Sequence 28, Appl
29	130.5	13.9	360	9 US-10-174-579-28	Sequence 28, Appl
30	130.5	13.9	360	9 US-10-174-582-28	Sequence 28, Appl
31	130.5	13.9	360	9 US-10-174-588-28	Sequence 28, Appl
32	130.5	13.9	360	9 US-10-175-739-28	Sequence 28, Appl
33	130.5	13.9	360	9 US-10-175-740-28	Sequence 28, Appl
34	130.5	13.9	360	9 US-10-175-743-28	Sequence 28, Appl
35	130.5	13.9	360	9 US-10-176-488-28	Sequence 28, Appl
36	130.5	13.9	360	9 US-10-176-492-28	Sequence 28, Appl
37	130.5	13.9	360	9 US-10-176-747-28	Sequence 28, Appl
38	130.5	13.9	360	9 US-10-176-750-28	Sequence 28, Appl
39	130.5	13.9	360	9 US-10-176-985-28	Sequence 28, Appl
40	130.5	13.9	360	9 US-10-176-987-28	Sequence 28, Appl
41	130.5	13.9	360	9 US-10-176-991-28	Sequence 28, Appl
42	130.5	13.9	360	9 US-10-176-992-28	Sequence 28, Appl
43	130.5	13.9	360	9 US-10-176-993-28	Sequence 28, Appl
44	130.5	13.9	360	9 US-10-184-658-28	Sequence 28, Appl
45	130.5	13.9	360	9 US-10-173-695-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-978-295A-52

; Sequence 52, Application US/09978295A
; Patent No. US2002015606A1

GENERAL INFORMATION:

APPLICANT: Ashkenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same
FILE REFERENCE: P2630P1C11
CURRENT APPLICATION NUMBER: US/09/978,295A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249

1	PRIOR FILING DATE: 1997-11-03	60/081817
2	PRIOR APPLICATION NUMBER: 60/065311	60/081817
3	PRIOR FILING DATE: 1997-11-13	60/081819
4	PRIOR APPLICATION NUMBER: 60/066364	60/081819
5	PRIOR FILING DATE: 1997-11-21	60/081952
6	PRIOR APPLICATION NUMBER: 60/077450	60/081952
7	PRIOR FILING DATE: 1998-03-10	60/081838
8	PRIOR APPLICATION NUMBER: 60/077632	60/081838
9	PRIOR FILING DATE: 1998-03-11	60/082568
10	PRIOR APPLICATION NUMBER: 60/077641	60/082568
11	PRIOR FILING DATE: 1998-03-11	60/082569
12	PRIOR APPLICATION NUMBER: 60/077649	60/082569
13	PRIOR FILING DATE: 1998-03-11	60/082704
14	PRIOR APPLICATION NUMBER: 60/077791	60/082704
15	PRIOR FILING DATE: 1998-03-12	60/082804
16	PRIOR APPLICATION NUMBER: 60/078004	60/082804
17	PRIOR FILING DATE: 1998-03-13	60/082700
18	PRIOR APPLICATION NUMBER: 60/078886	60/082700
19	PRIOR FILING DATE: 1998-03-20	60/082799
20	PRIOR APPLICATION NUMBER: 60/078936	60/082799
21	PRIOR FILING DATE: 1998-03-20	60/082796
22	PRIOR APPLICATION NUMBER: 60/078910	60/082796
23	PRIOR FILING DATE: 1998-03-20	60/083336
24	PRIOR APPLICATION NUMBER: 60/078939	60/083336
25	PRIOR FILING DATE: 1998-03-20	60/083322
26	PRIOR APPLICATION NUMBER: 60/079294	60/083322
27	PRIOR FILING DATE: 1998-03-25	60/083392
28	PRIOR APPLICATION NUMBER: 60/079656	60/083392
29	PRIOR FILING DATE: 1998-03-26	60/083495
30	PRIOR APPLICATION NUMBER: 60/079664	60/083495
31	PRIOR FILING DATE: 1998-03-27	60/083496
32	PRIOR APPLICATION NUMBER: 60/079689	60/083496
33	PRIOR FILING DATE: 1998-03-27	60/083499
34	PRIOR APPLICATION NUMBER: 60/079663	60/083499
35	PRIOR FILING DATE: 1998-03-27	60/083545
36	PRIOR APPLICATION NUMBER: 60/079728	60/083545
37	PRIOR FILING DATE: 1998-03-27	60/083554
38	PRIOR APPLICATION NUMBER: 60/079786	60/083554
39	PRIOR FILING DATE: 1998-03-27	60/083558
40	PRIOR APPLICATION NUMBER: 60/079920	60/083558
41	PRIOR FILING DATE: 1998-03-30	60/083559
42	PRIOR APPLICATION NUMBER: 60/079923	60/083559
43	PRIOR FILING DATE: 1998-03-30	60/083500
44	PRIOR APPLICATION NUMBER: 60/080105	60/083500
45	PRIOR FILING DATE: 1998-03-31	60/083742
46	PRIOR APPLICATION NUMBER: 60/080107	60/083742
47	PRIOR FILING DATE: 1998-03-31	60/084366
48	PRIOR APPLICATION NUMBER: 60/080165	60/084366
49	PRIOR FILING DATE: 1998-03-31	60/084414
50	PRIOR APPLICATION NUMBER: 60/080194	60/084414
51	PRIOR FILING DATE: 1998-03-31	60/084441
52	PRIOR APPLICATION NUMBER: 60/080327	60/084441
53	PRIOR FILING DATE: 1998-04-01	60/084637
54	PRIOR APPLICATION NUMBER: 60/080328	60/084637
55	PRIOR FILING DATE: 1998-04-01	60/084639
56	PRIOR APPLICATION NUMBER: 60/080333	60/084639
57	PRIOR FILING DATE: 1998-04-01	60/084640
58	PRIOR APPLICATION NUMBER: 60/080334	60/084640
59	PRIOR FILING DATE: 1998-04-01	60/084598
60	PRIOR APPLICATION NUMBER: 60/081070	60/084598
61	PRIOR FILING DATE: 1998-04-08	60/084600
62	PRIOR APPLICATION NUMBER: 60/081049	60/084600
63	PRIOR FILING DATE: 1998-04-08	60/084627
64	PRIOR APPLICATION NUMBER: 60/081071	60/084627
65	PRIOR FILING DATE: 1998-04-08	60/084643
66	PRIOR APPLICATION NUMBER: 60/081195	60/084643
67	PRIOR FILING DATE: 1998-04-08	60/085339
68	PRIOR APPLICATION NUMBER: 60/081203	60/085339
69	PRIOR FILING DATE: 1998-04-09	60/085338
70	PRIOR APPLICATION NUMBER: 60/081229	60/085338
71	PRIOR FILING DATE: 1998-04-09	60/085323
72	PRIOR APPLICATION NUMBER: 60/081955	60/085323
73	PRIOR FILING DATE: 1998-04-15	60/085582
74	PRIOR APPLICATION NUMBER: 60/085582	60/085582

;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085689
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 92.9%; Score 870; DB 9; Length 321;
Best Local Similarity 64.7%; Pred. No. 1.le-66;
Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

Qy 1 MGILLGLLLGLHVTDTYGRPILEVPEVSGTGPWKGDVNLPCYDPLQGYTVLVKWLVR 60
Db 1 MGILLGLLLGLHVTDTYGRPILEVPEVSGTGPWKGDVNLPCYDPLQGYTVLVKWLVR 60
Qy 61 GSDPTVIFLRSSGDHIOQAKYQGRHLVSHKVPDVSLSQLSTLEMDRSHYTCVETWQTP 120
Db 61 GSDPTVIFLRSSGDHIOQAKYQGRHLVSHKVPDVSLSQLSTLEMDRSHYTCVETWQTP 120
Qy 121 DGNQVVRDKITELRVQ----- 136
Db 121 DGNQVVRDKITELRVQKLSVSKPTVTGSGYGTYPQGMRIISLQCAQSPISYIWKQ 180
Qy 137 -----KHSSKLLTKK 146
Db 181 OTNNOEPIKATLSLLFKPAVIADSGSYFCTAKQGVGEHQSDIVKFWVKDSSKLLTKK 240
Qy 147 TEAPTMTYPLKATSTVKOSWDWTTDMDG 175
Db 241 TEAPTMTYPLKATSTVKOSWDWTTDMDG 269

RESULT 2
US-09-763-902b-6

;; Sequence 52, Application US/09978697
;; Patent No. US20020169284A1

;; GENERAL INFORMATION:

;; APPLICANT: Ashkenazi, Avi
;; APPLICANT: Baker Kevin P.
;; APPLICANT: Botstein, David
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Eaton, Dan
;; APPLICANT: Ferrara, Napoleon
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Fong, Sherman
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerber, Hanspeter
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Hillan, Kenneth J.
;; APPLICANT: Kijavini, Ivar J.
;; APPLICANT: Kuo, Sophia S.
;; APPLICANT: Napier, Mary A.
;; APPLICANT: Pan, James
;; APPLICANT: Paoni, Nicholas F.
;; APPLICANT: Roy, Margaret Ann
;; APPLICANT: Shelton, David L.
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Williams, P. Mickey
;; APPLICANT: Wood, William I.

;; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

;; TITLE OF INVENTION: Acids Encoding the Same
;; FILE REFERENCE: P2630P1C27
;; CURRENT APPLICATION NUMBER: US/09/978,697
;; CURRENT FILING DATE: 2001-10-16
;; PRIOR APPLICATION NUMBER: 09/918585
;; PRIOR FILING DATE: 2001-07-30
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/064249
;; PRIOR FILING DATE: 1997-11-03
;; PRIOR APPLICATION NUMBER: 60/065311
;; PRIOR FILING DATE: 1997-11-13
;; PRIOR APPLICATION NUMBER: 60/066364
;; PRIOR FILING DATE: 1997-11-21
;; PRIOR APPLICATION NUMBER: 60/077450
;; PRIOR FILING DATE: 1998-03-10
;; PRIOR APPLICATION NUMBER: 60/077632
;; PRIOR FILING DATE: 1998-03-11
;; PRIOR APPLICATION NUMBER: 60/077641
;; PRIOR FILING DATE: 1998-03-11
;; PRIOR APPLICATION NUMBER: 60/077649
;; PRIOR FILING DATE: 1998-03-11
;; PRIOR APPLICATION NUMBER: 60/077791
;; PRIOR FILING DATE: 1998-03-12
;; PRIOR APPLICATION NUMBER: 60/078004
;; PRIOR FILING DATE: 1998-03-13
;; PRIOR APPLICATION NUMBER: 60/078886
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/078936
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/078910
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/078939
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/079294
;; PRIOR FILING DATE: 1998-03-25
;; PRIOR APPLICATION NUMBER: 60/079656
;; PRIOR FILING DATE: 1998-03-26
;; PRIOR APPLICATION NUMBER: 60/079664
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079663
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079728
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079786
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079920
;; PRIOR FILING DATE: 1998-03-30
;; PRIOR APPLICATION NUMBER: 60/079923
;; PRIOR FILING DATE: 1998-03-30
;; PRIOR APPLICATION NUMBER: 60/080105
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080107
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080165
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080194
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080327
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080328
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080333
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080334
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/081070
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081049
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081071

;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081195
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081203
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081229
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081955
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081817
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081819
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081952
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081838
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/082568
;; PRIOR FILING DATE: 1998-04-21
;; PRIOR APPLICATION NUMBER: 60/082569
;; PRIOR FILING DATE: 1998-04-21
;; PRIOR APPLICATION NUMBER: 60/082704
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082804
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082700
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082797
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082796
;; PRIOR FILING DATE: 1998-04-23
;; PRIOR APPLICATION NUMBER: 60/083336
;; PRIOR FILING DATE: 1998-04-27
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/083392
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083495
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083496
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083499
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083545
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083554
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083558
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083559
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083500
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083742
;; PRIOR FILING DATE: 1998-04-30
;; PRIOR APPLICATION NUMBER: 60/084366
;; PRIOR FILING DATE: 1998-05-05
;; PRIOR APPLICATION NUMBER: 60/084414
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084441
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084639
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084640
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084598
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084600
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084627
;; PRIOR FILING DATE: 1998-05-07

;; PRIOR APPLICATION NUMBER: 60/084643
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085582
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085689
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697
;; PRIOR APPLICATION NUMBER: 60/085697
Query Match 92.9%; Score 870; DB 9; Length 321;
Best Local Similarity 64.7%; Pred. No. 1,1e-66;
Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;
QY 1 MGILLGILLGHLTVDTYGRPILEVPESVTGPKGVDVNLCTYDPLOGYTQVLYKWLVR 60
DB 1 MGILLGILLGHLTVDTYGRPILEVPESVTGPKGVDVNLCTYDPLOGYTQVLYKWLVR 60
QY 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSLSQSTLEMDRSHYTCVTTWQTP 120
DB 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSLSQSTLEMDRSHYTCVTTWQTP 120
QY 121 DGNQVVRDKITELRVQ----- 136
DB 121 DGNQVVRDKITELRVQKLSVSKPTVTGSGYGVFTVPGQMRISLQCCARGSPPIYIWKQ 180
QY 137 -----KHSKLLKTK 146
DB 181 QTNQPEPIKVAITLSTLLFKPAVIADSSYFCTAKGVGSEQHSDIVKRVKDSKLLKTK 240
QY 147 TEAPTTMTYPLKATSTYKQSDWDTTMDG 175
DB 241 TEAPTTMTYPLKATSTYKQSDWDTTMDG 269

RESULT 3

US-09-978-192A-52
;; Sequence 52, Application US/09978192A.
;; Patent No. US20020177553A1
;; GENERAL INFORMATION:
;; APPLICANT: Ashkenazi, Avi
;; APPLICANT: Baker Kevin P.
;; APPLICANT: Botstein, David
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Eaton, Dan
;; APPLICANT: Ferrara, Napoleon
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Fong, Sherman
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerber, Hanspeter
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Hillan, Kenneth J.
;; APPLICANT: Kijavini, Ivar J.
;; APPLICANT: Kuo, Sophia S.
;; APPLICANT: Napier, Mary A.

APPLICANT: Pan, James;
APPLICANT: Peoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630PIC9
CURRENT APPLICATION NUMBER: US/09/978,192A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918595
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062230
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080155
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639

FILE REFERENCE: P1216R1(US)
CURRENT APPLICATION NUMBER: US/09/953,499
CURRENT FILING DATE: 2001-09-14
PRIOR APPLICATION NUMBER: US/09/254,465
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: PCT/US98/24855
PRIOR FILING DATE: 1998-11-20
PRIOR APPLICATION NUMBER: US 60/066,364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 60/078,936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: PCT/US98/19437
PRIOR FILING DATE: 1998-09-17
NUMBER OF SEQ ID NOS: 30
SEQ ID NO 2
LENGTH: 321
TYPE: PRT
ORGANISM: Homo sapiens
US-09-953-499-2

Query Match 92.9%; Score 870; DB 9; Length 321;
Best Local Similarity 64.7%; Pred. No. 1.le-66;
Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;
Qy 1 MGILLG LLLGHLTVDTYGRPILEVPESTGPKGDVNLPCYDPLQGYTVLVKLVQR 60
Db 1 MGILLG LLLGHLTVDTYGRPILEVPESTGPKGDVNLPCYDPLQGYTVLVKLVQR 60
Qy 61 GSDPVTIFLRDSSGDHIQQAQYGRHSHVSKVPGDVSQSLSTLEMDRSHYTCVETWTP 120
Db 61 GSDPVTIFLRDSSGDHIQQAQYGRHSHVSKVPGDVSQSLSTLEMDRSHYTCVETWTP 120
Qy 121 DGNQVVRDKITELRVQ----- 136
Db 121 DGNQVVRDKITELRVQKLSVSKPTVTGSGYFTVPGQMRISLQOARGSPPIISYWKQ 180
Qy 137 -----KHSSKLLKTK 146
Db 181 QTNQEPKIVATLSTLLFKPAVIADSGSYFCTAKQGVSEQHSQDIVKFWKDSKLLKTK 240
Qy 147 TEAPTTMTYPLKATSVKQSWDWTMDMG 175
Db 241 TEAPTTMTYPLKATSVKQSWDWTMDMG 269
RESULT 4
US-09-953-499-2
Sequence 2, Application US/09953499
Publication No.: US20020182206A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Fong, Sherman
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Napier, Mary A.
APPLICANT: Tamas, Daniel
APPLICANT: Wood, William I.
TITLE OF INVENTION: COMPOUNDS, COMPOSITIONS AND METHODS FOR THE TREATMENT
TITLE OF INVENTION: OF DISEASES CHARACTERIZED BY A33- RELATED ANTIGENS

Query Match 92.9%; Score 870; DB 9; Length 321;
Best Local Similarity 64.7%; Pred. No. 1.le-66;
Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;
Qy 1 MGILLG LLLGHLTVDTYGRPILEVPESTGPKGDVNLPCYDPLQGYTVLVKLVQR 60
Db 1 MGILLG LLLGHLTVDTYGRPILEVPESTGPKGDVNLPCYDPLQGYTVLVKLVQR 60
Qy 61 GSDPVTIFLRDSSGDHIQQAQYGRHSHVSKVPGDVSQSLSTLEMDRSHYTCVETWTP 120
Db 61 GSDPVTIFLRDSSGDHIQQAQYGRHSHVSKVPGDVSQSLSTLEMDRSHYTCVETWTP 120
Qy 121 DGNQVVRDKITELRVQ----- 136
Db 121 DGNQVVRDKITELRVQKLSVSKPTVTGSGYFTVPGQMRISLQOARGSPPIISYWKQ 180
Qy 137 -----KHSSKLLKTK 146
Db 181 QTNQEPKIVATLSTLLFKPAVIADSGSYFCTAKQGVSEQHSQDIVKFWKDSKLLKTK 240
Qy 147 TEAPTTMTYPLKATSVKQSWDWTMDMG 175
Db 241 TEAPTTMTYPLKATSVKQSWDWTMDMG 269
RESULT 5
US-09-999-832A-52
Sequence 52, Application US/09999832A
Publication No.: US20020192706A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Pao, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.

APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secretds and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630PIC63
CURRENT APPLICATION NUMBER: US/09/999,832A
CURRENT FILING DATE: 2001-10-24
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598

; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

Query Match 92.9%; Score 870; DB 9; Length 321;

Best Local Similarity 64.7%; Pred. No. 1.le-66;
Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

QY 1 MGIILGLLGLHLVDTGRPILEVPESTGPKGDVNLPCYDPLQGYTVLVKWLVR 60
DB 1 MGIILGLLGLHLVDTGRPILEVPESTGPKGDVNLPCYDPLQGYTVLVKWLVR 60
QY 61 GSDPTVIFLRSSGDHIOQAKYQGRHLYSHKVPDVSLSQSLSTLEMDRSHYTCVTTWTP 120
DB 61 GSDPTVIFLRSSGDHIOQAKYQGRHLYSHKVPDVSLSQSLSTLEMDRSHYTCVTTWTP 120
QY 121 DGNQVVRDKITELRVQ----- 136
DB 121 DGNQVVRDKITELRVQKLSVSKPTVTTCGSGYGTVPQGMRIISLOQARGSPPISYWKQ 180
QY 137 -----KHSSKLLKTK 146
DB 181 QTNNOEPKIVATLSLLFPKPAVIAIDSGSVFCTAKQGVSEHQSDIVKFFVVDSSKLLKTK 240
QY 147 TEAPTTMYPLKATSVKQSWDWTMTDMDG 175
DB 241 TEAPTTMYPLKATSVKQSWDWTMTDMDG 269

RESULT 6

US-09-978-189-52
; Sequence 52, Application us/09978189
; Publication No. US20030004102a1
; GENERAL INFORMATION:

; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey J.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher

; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2630P1C7
; CURRENT APPLICATION NUMBER: US/09/978,189
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194

;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080327
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080328
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080333
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080334
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/081070
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081049
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081071
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081195
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081203
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081229
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081955
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081817
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081819
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081952
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081838
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/082568
;; PRIOR FILING DATE: 1998-04-21
;; PRIOR APPLICATION NUMBER: 60/082569
;; PRIOR FILING DATE: 1998-04-21
;; PRIOR APPLICATION NUMBER: 60/082704
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082804
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082700
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082797
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082796
;; PRIOR FILING DATE: 1998-04-23
;; PRIOR APPLICATION NUMBER: 60/083336
;; PRIOR FILING DATE: 1998-04-27
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/083392
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083495
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083496
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083499
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083545
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083554
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083558
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083559
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083500
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083742
;; PRIOR FILING DATE: 1998-04-30
;; PRIOR APPLICATION NUMBER: 60/084366
;; PRIOR FILING DATE: 1998-05-05
;; PRIOR APPLICATION NUMBER: 60/084414
;; PRIOR FILING DATE: 1998-05-06

;; PRIOR APPLICATION NUMBER: 60/084441
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084639
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084640
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084598
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084600
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084627
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084643
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085582
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085689
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 92.9%; Score 870; DB 9; Length 321;

Best Local Similarity 84.7%; Pred. No. 1.le-66;

Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

QY 1 MGILLGLLLGLHGLVDTYGRPILEVPSVTPGPKGVDNLPCTYLOGYTOVLVKLVQR 60
DB 1 MGILLGLLLGLHGLVDTYGRPILEVPSVTPGPKGVDNLPCTYLOGYTOVLVKLVQR 60
QY 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSLSQSTLEMDRSHYTCVETWQTP 120
DB 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSLSQSTLEMDRSHYTCVETWQTP 120
QY 121 DGNQVVRDKITELRVQ----- 136
DB 121 DGNQVVRDKITELRVQKLSVSKPTVTGSGYGTVPQGMRLSQOARGSPPIYWKQ 180
QY 137-----KHSSKLLKTK 146
DB 181 QTNQOEPIKIVATLSTLLFKPAVIADSGSYFTAKQGVGSEQHSDIVKPVVKDSSKLLKTK 240
QY 147 TEAPTTMTYPLKATSTYKQSDWTTDMDG 175
DB 241 TEAPTTMTYPLKATSTYKQSDWTTDMDG 269

RESULT 7

US-09-905-291A-213
; Sequence 213, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905/291A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 213
LENGTH: 360
TYPE: PRT
ORGANISM: Homo sapiens
US-09-905-291A-213

Query Match 13.9%; Score 130.5; DB 9; Length 360;
Best Local Similarity 28.4%; Pred. No. 0.00086;
Matches 52; Conservative 20; Mismatches 42; Indels 69; Gaps 10;
QY 1 MGILLGLLLGLHVTVDYGRP-----ILEVPESVT 30
DB 1 MGILLGLVPL--LLPGSYGLPTNGEYNSANDONLGNHGKDLLNGVKLVVETPEEL 58

QY 31 GPMKG-DVNLPC--TYDP-LQGYTOVLVKW--LVQGS DP-----VTIFLRDSS-GDHIQ 79
DB 59 FTQOGASVILPCRYRIEPALVSPRRVRKWKWKLSENAPKQVLAIGLHRHSGD--- 114
QY 80 AKYQGRHVS HKVPGDYS LQSLSTLEMDRSHYTCV-----TWQT 119
DB 115 --YGRVHLRQDKHEHDYSLEIQDLRLDYGRYRCEVIDGLESGLELELRGVVFPYQS 172
QY 120 PDG 122
DB 173 PNG 175

RESULT 8
US-09-902-853-213
Sequence 213, Application US/09902853
Publication No. US20020192659A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/902,853
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US/09/665,350
PRIOR FILING DATE: 2000-09-16
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02

```

; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 213
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-907-824-213

Query Match          13.9%   Score 130.5; DB 9; Length 360;
Best Local Similarity 28.4%; Pred. No. 0.00086;
Matches 52; Conservative 20; Mismatches 42; Indels 69; Gaps 10;

QY    1 MGILLGULLLGLHTVDYGRP-----ILEVPSVSPT 30
      |||::||| | |::|||
Db     1 MGILLLVPL--LLPGSYGLPFYNGFYNSANDNLGNHKGKDLLNKGVLVETPEETL 58
      ::||| | | | | |::|||

QY    31 GPWK-G-DYNLPC--TYDP-LQGYTQLVKW--LVQRSDP----VTFLRDS--GDHIQQ 79
      ::||| | | | | |::|||
Db     59 FTYGASVILPCKRYREPALVPSPRRVRKMKLSENGAPEKDVLVAIGLRHSFGD--- 114
      ::||| | | | | |::|||

QY    80 AKYQGRCHSVHKVPGDVYSLOLSLEMDRRSHYTCEV-----TWQT 119
      |||::||| | |::|||
Db     115 --YQGRVHLRQDKHDVSLEIQDLRLDEDYGRYRCVEIDGLESGLVELELRGVVFPIQS 172
      ||| | | | | |::|||

QY    120 PDG 122
      |||
Db     173 PNG 175

RESULT 10
US-09-907-841-213
; Sequence 213, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc

```


APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secretd and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,841
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 213
LENGTH: 360
TYPE: PRT
ORGANISM: Homo sapiens
US-09-907-841-213

Query Match 13.9%; Score 130.5; DB 9; Length 360;
Best Local Similarity 28.4%; Pred. No. 0.00086;
Matches 52; Conservative 20; Mismatches 42; Indels 69; Gaps 10;

QY 1 MGILLGLLGLHGLTVTYGRP-----ILEVPESVT 30
DB 1 MGILLVPLL--LLPGSYGLPFTNGFYNSANDQNLGNHGKDLNGVKLVVETPEETL 58
QY 31 GPWKG-DVNLPC--TYDP-LQGYTOVLVKW--LVQRGSDP-----VTFILRDS--GDHIQ 79
DB 59 FTYGASVILPCRYRVEPALVSPRRVKKWKLSENGAPEKDVLAIGLHRSGD---- 114
QY 80 KYQGRHLVSHKVPQGVSLQSLTLEMDRSHYCEV-----TWQT 119
DB 115 --YQGRVHLQRKDEHVSLEIQDLRLDYGRCYCEVDIGLEBSGLVLELGRGVFPYQS 172
QY 120 PDG 122
DB 173 PNG 175

RESULT 11
US-09-904-011-213
Sequence 213, Application US/09904011
Publication No. US20030003530A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secretd and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/904,011
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05

```
QY      1 MGILLGLLHLGLTVDVTCGRP-----ILEVPESVT 30
         ||:||:|||||
Db       1 MGLLLLVLPL--LLPGSYGLPYNGFYYSANSANDQNGLNGCHKDLLNGKLVVEPEETL 58
         ||:||:|||||

QY      31 GPWK-GDVNLP-C-TVDP-LQGTVOLVKW--LVQRGSDP----VTIFLRDSS-GBHIOQ 79
         :|||:|||:|||:|||:|||
Db       59 FTYOGASVILPCRYRYEPALVSRRVRVKWKLSENGAPEKDVLVAIGLRHSFGD--- 114
         :|||:|||:|||:|||:|||

QY      80 AKYQGRLHVSHRKPVGDSVLQLSTLEMDDRSHTCEV-----TWOT 119
         |||||::|||::|||::|||
Db       115 --YQGRHLRQKHDEHDVLSIEIQDLRDLYGRYRCVEDGLESGVLVELRGVVFPYGOS 172
         |||||::|||::|||::|||
```

```

: GENERAL INFORMATION:
:
: APPLICANT: Baker, Kevin P.
:
: APPLICANT: Chen, Jian
:
: APPLICANT: Desnoyers, Luc
:
: APPLICANT: Goddard, Audrey
:
: APPLICANT: Godowski, Paul J.
:
: APPLICANT: Gurney, Austin L.
:
: APPLICANT: Pan, James
:
: APPLICANT: Smith, Victoria
:
: APPLICANT: Watanabe, Colin K.
:
: APPLICANT: Wood, William I.
:
: APPLICANT: Zhang, Zemin
:
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
:

```

```
FILE REFERENCE: P3430R1C50
CURRENT APPLICATION NUMBER: US/10/175,737
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 28
LENGTH: 360
TYPE: PRT
ORGANISM: Homo Sapien
US-10-175-737-28

Query Match      13.9%; Score 130.5; DB 9; Length 360;
Best Local Similarity 28.4%; Pred. No. 0.00086;
Matches 52; Conservative 20; Mismatches 42; Indels 69; Gaps 10;

QY 1 MGILGLLLGLHGLTVDYGRP-----IILEVPESVT 30
DB 1 MGILGLLVPL--LLPGSYGLPFYNGFYNSANDQNLNGHGKDLLNGVKLVVETPEETL 58
QY 31 GPWKG-DVNLPC--TYDP-LOGYTOVLVKW--LVQRGSDP-----VTIFLRDSS-GDHIQ 79
DB 59 FTYQASVILPCRYEPALVSPRRVRVKKWKLSENGAPEKDVLAIGLRHRSFGD---- 114
QY 80 AKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCV-----TWOT 119
DB 115 --YQGRVHLRQDKHDSVLSIQDLRLDYGRYRCEVIDGLEDESGLVELELRGVVFPYQS 172
QY 120 PDG 122
DB 173 PNG 175

RESULT 15
US-09-763-902b-6.rapb
Sequence 213, Application US/09906742
Publication No: US20030023054A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT FILING DATE: 2001-07-16
PRIOR APPLICATION NUMBER: US/09/906,742
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
```

```
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 213
LENGTH: 360
TYPE: PRT
ORGANISM: Homo Sapien
US-09-906-742-213

Query Match      13.9%; Score 130.5; DB 9; Length 360;
Best Local Similarity 28.4%; Pred. No. 0.00086;
Matches 52; Conservative 20; Mismatches 42; Indels 69; Gaps 10;

QY 1 MGILGLLLGLHGLTVDYGRP-----IILEVPESVT 30
DB 1 MGILGLLVPL--LLPGSYGLPFYNGFYNSANDQNLNGHGKDLLNGVKLVVETPEETL 58
QY 31 GPWKG-DVNLPC--TYDP-LOGYTOVLVKW--LVQRGSDP-----VTIFLRDSS-GDHIQ 79
DB 59 FTYQASVILPCRYEPALVSPRRVRVKKWKLSENGAPEKDVLAIGLRHRSFGD---- 114
QY 80 AKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCV-----TWOT 119
DB 115 --YQGRVHLRQDKHDSVLSIQDLRLDYGRYRCEVIDGLEDESGLVELELRGVVFPYQS 172
QY 120 PDG 122
DB 173 PNG 175

Search completed: March 4, 2003, 15:07:26
Job time : 6.18902 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:32 ; Search time 49.2588 Seconds
(without alignments)
3007.543 Million cell updates/sec

Title: US-09-763-902b-5

Perfect score: 3638

Sequence: 1 MSVDKAEKCGSLLTWLTQFH.....SRGPLGRLASINLRPTDKH 719

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3631	99.8	719	4	O60562
2	3611	99.3	717	4	Q96ED9
3	3102.5	85.3	716	11	Q8VCN4
4	2113.5	58.1	490	11	Q9LJL2
5	2048.5	56.3	473	11	Q8R347
6	1758	48.3	718	4	Q9BRL3
7	1667	45.8	728	4	O60561
8	864	23.7	678	5	O61493
9	854	23.5	679	5	Q9VJ15
10	850	23.4	679	5	Q24185
11	593.5	16.3	256	11	Q9CTN6
12	385.5	10.6	709	4	O14997
13	385.5	10.6	742	4	Q9UI01
14	378	10.4	1235	5	Q8SX64
15	378	10.4	1381	5	Q9VZT7
16	342.5	9.4	2442	4	O14812

ALIGNMENTS

RESULT 1

O60562 PRELIMINARY; PRT: 719 AA.

AC O60562;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hook2 protein.
GN HOOK2.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99126435; PubMed=9927460;
RA Kramer H., Phistry M.;
RT "Genetic analysis of hook, a gene required for endocytic trafficking
in Drosophila."
RL Genetics 151:675-684(1999).
DR EMBL; AF044924; AAC09299.1;
SQ SEQUENCE 719 AA; 83163 MW; 2E14C6B1435CE96C CRC64;

Query Match 99.8%; Score 3631; DB 4; Length 719;
Best Local Similarity 99.9%; Pred. No. 1.1e-163;
Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSVDKAEKCGSLLTWLTQFHVPSPCASQDLSGLAVAYVNLQIDPSWNEAWLQIGSED 60
|||||
DB 1 MSVDKAEKCGSLLTWLTQFHVPSPCASQDLSGLAVAYVNLQIDPSWNEAWLQIGSED 60
|||||

QY 61 PGPNWLKYSNKLKWLRSLEVESQDVLAPVSEHLPDVSLSLIGFSDPAELKGLQLVLG 120
|||||

DB 61 PGPNWLKYSNKLKWLRSLEVESQDVLAPVSEHLPDVSLSLIGFSDPAELKGLQLVLG 120
|||||

QY 121 CAISCEKKQDHIQRIIMTLEESVQHVYVMEAIQELMTKDTPDLSLSPETYGNFDSQSRYYFL 180
|||||

DB 121 CAISCEKKQDHIQRIIMTLEESVQHVYVMEAIQELMTKDTPDLSLSPETYGNFDSQSRYYFL 180
|||||

Q9P2M7 homo sapien
O60588 homo sapien
Q9H450 homo sapien
Q8TX14 methanopyru
Q23529 caenorhabdi
Q25893 plasmodium
Q9BQS8 homo sapien
Q26080 placoceten
Q63731 rattus norv
Q9HJ22 aequipeten
Q9HJ21 aequipeten
Q17042 aequipeten
Q70365 mus musculu
Q9B216 homo sapien
O43241 homo sapien
Q8WZA3 homo sapien
Q14980 homo sapien
Q75033 homo sapien
Q9H430 homo sapien
Q8TDA9 homo sapien
Q9HKT9 leishmania
Q9TY21 drosophila
Q20641 caenorhabdi
Q9PTD7 xenopus lae
Q9U0S5 mytilus gal
Q9HJ23 aequipeten
Q9U0S7 mytilus gal

```
QY 181 SEAEDEGDELQORCLDLERQRLMLLSEKQSLAQENAGNLRMRGPEGEGTGLTAKKLL 240
DB 181 SEAEDEGDELQORCLDLERQRLMLLSEKQSLAQENAGNLRMRGPEGEGTGLTAKKLL 240
QY 241 LOSOLEQOEENFLESREDELRCAELEREVAELQHRNOALTSLAQEAQALKDEMDL 300
DB 241 LOSOLEQOEENFLESREDELRCAELEREVAELQHRNOALTSLAQEAQALKDEMDL 300
QY 301 ROSSERAGQLEATLTSRRRLGELRELRRQVROLEERNAGHAERTROLEDELRRAGSLRA 360
DB 301 ROSSERAGQLEATLTSRRRLGELRELRRQVROLEERNAGHAERTROLEDELRRAGSLRA 360
QY 361 OLEAQRROVQELQORQOEAMKAELKWLFECHNLEKYEVSVKERLLAERDSLEANE 420
DB 361 OLEAQRROVQELQORQOEAMKAELKWLFECHNLEKYEVSVKERLLAERDSLEANE 420
QY 421 LRCALQPRGLTQADPSLDPTSTPVDNLAAELPAELRETLRLQLENKRLCRQEAADRE 480
DB 421 LRCALQPRGLTQADPSLDPTSTPVDNLAAELPAELRETLRLQLENKRLCRQEAADRE 480
QY 481 ROELQRLHLEDANRARGHLETOHRLNQOQLSELRAQVEDLQALQEOGGKTEDAISILK 540
DB 481 ROELQRLHLEDANRARGHLETOHRLNQOQLSELRAQVEDLQALQEOGGKTEDAISILK 540
QY 541 RKLEHLQKLHLEADLELQKREYIEELEPPDTSSTARRIEBELQHLNQLKDDADLRAMEERY 600
DB 541 RKLEHLQKLHLEADLELQKREYIEELEPPDTSSTARRIEBELQHLNQLKDDADLRAMEERY 600
QY 601 RRYVDKARWQMEPKORPAAGAPPELHSLRTOLRERDVIRHLEMDFEKRSQREOEE 660
DB 601 RRYVDKARWQMEPKORPAAGAPPELHSLRTOLRERDVIRHLEMDFEKRSQREOEE 660
QY 661 KLLISAWYNGMALQORAGEERAPAHQSFLLAQORLATNSRRGPLGRSLASLNLRLPTDKH 719
DB 661 KLLISAWYNGMALQORAGEERAPAHQSFLLAQORLATNSRRGPLGRSLASLNLRLPTDKH 719
```

RESULT 2

```
Q96ED9 PRELIMINARY; PRT; 717 AA.
ID Q96ED9 AC Q96ED9
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Unknown (protein for MGC:8882).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012443; AAH12443.1;
KW SEQUENCE 717 AA; 83013 MW; A00810F65EB4C98F CRC64;
```

Query Match 99.3%; Score 3611; DB 4; Length 717;
Best Local Similarity 99.6%; Pred. No. 9.4e-163;
Matches 716; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

```
QY 1 MSVDKAEICGSLTTLWQTFHVPSPCASQDLSGLAVAYVLNQIDPSWFWNEAWLQGISD 60
DB 1 MSVDKAEICGSLTTLWQTFHVPSPCASQDLSGLAVAYVLNQIDPSWFWNEAWLQGISD 60
QY 61 PGPWKLKVNKLKWLRLSLVEYSQDVLAHPVSEELHPDVSLIGESDPAELGKLLQLVLG 120
DB 61 PGPWKLKVNKLKWLRLSLVEYSQDVLAHPVSEELHPDVSLIGESDPAELGKLLQLVLG 120
QY 121 CAISCEKKODHIQRIWTLTEESVQHVVMIAIQELMTKDPDLSLSPETYNFQDSQSRYYFL 180
DB 121 CAISCEKKODHIQRIWTLTEESVQHVVMIAIQELMTKDPDLSLSPETYNFQDSQSRYYFL 180
```

```
QY 181 SEAEDEGDELQORCLDLERQRLMLLSEKQSLAQENAGNLRMRGPEGEGTGLTAKKLL 240
DB 181 SEAEDEGDELQORCLDLERQRLMLLSEKQSLAQENAGNLRMRGPEGEGTGLTAKKLL 240
QY 241 LOSOLEQOEENFLESREDELRCAELEREVAELQHRNOALTSLAQEAQALKDEMDL 300
DB 241 LOSOLEQOEENFLESREDELRCAELEREVAELQHRNOALTSLAQEAQALKDEMDL 300
QY 301 ROSSERAGQLEATLTSRRRLGELRELRRQVROLEERNAGHAERTROLEDELRRAGSLRA 360
DB 301 ROSSERAGQLEATLTSRRRLGELRELRRQVROLEERNAGHAERTROLEDELRRAGSLRA 360
QY 361 OLEAQRROVQELQORQOEAMKAELKWLFECHNLEKYEVSVKERLLAERDSLEANE 420
DB 361 OLEAQRROVQELQORQOEAMKAELKWLFECHNLEKYEVSVKERLLAERDSLEANE 420
QY 421 LRCALQPRGLTQADPSLDPTSTPVDNLAAELPAELRETLRLQLENKRLCRQEAADRE 480
DB 421 LRCALQPRGLTQADPSLDPTSTPVDNLAAELPAELRETLRLQLENKRLCRQEAADRE 480
QY 481 ROELQRLHLEDANRARGHLETOHRLNQOQLSELRAQVEDLQALQEOGGKTEDAISILK 540
DB 481 ROELQRLHLEDANRARGHLETOHRLNQOQLSELRAQVEDLQALQEOGGKTEDAISILK 538
QY 541 RKLEHLQKLHLEADLELQKREYIEELEPPDTSSTARRIEBELQHLNQLKDDADLRAMEERY 600
DB 539 RKLEHLQKLHLEADLELQKREYIEELEPPDTSSTARRIEBELQHLNQLKDDADLRAMEERY 598
QY 601 RRYVDKARWQMEPKORPAAGAPPELHSLRTOLRERDVIRHLEMDFEKRSQREOEE 660
DB 599 RRYVDKARWQMEPKORPAAGAPPELHSLRTOLRERDVIRHLEMDFEKRSQREOEE 658
QY 661 KLLISAWYNGMALQORAGEERAPAHQSFLLAQORLATNSRRGPLGRSLASLNLRLPTDKH 719
DB 659 KLLISAWYNGMALQORAGEERAPAHQSFLLAQORLATNSRRGPLGRSLASLNLRLPTDKH 717
```

RESULT 3

```
Q8VCN4 PRELIMINARY; PRT; 716 AA.
ID Q8VCN4 AC Q8VCN4
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical 83.4 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019486; AAH19486.1;
KW Hypothetical protein.
SQ SEQUENCE 716 AA; 83366 MW; 313333ABF8C6AA50 CRC64;
```

Query Match 85.3%; Score 3102.5; DB 11; Length 716;
Best Local Similarity 85.7%; Pred. No. 8.4e-139;
Matches 616; Conservative 47; Mismatches 53; Indels 3; Gaps 2;

```
QY 1 MSVDKAEICGSLTTLWQTFHVPSPCASQDLSGLAVAYVLNQIDPSWFWNEAWLQGISD 60
DB 1 MSVDKAEICGSLTTLWQTFHVPSPCASQDLSGLAVAYVLNQIDPSWFWNEAWLQGISD 60
QY 61 PGPWKLKVNKLKWLRLSLVEYSQDVLAHPVSEELHPDVSLIGESDPAELGKLLQLVLG 120
DB 61 SPSWRLKVRKLEIKLQSLVEYSQDVLAHPVSDQHPVSLIGESDPAELGKLLQLVLG 120
QY 121 CAISCEKKODHIQRIWTLTEESVQHVVMIAIQELMTKDPDLSLSPETYNFQDSQSRYYFL 180
DB 121 CAISCEKKODHIQRIWTLTEESVQHVVMIAIQELMTKDPDLSLSPETYNFQDSQSRYYFL 180
```

QY	411	ROSLSREANEELRCAQLQPGRLGTLQADPSLDPTSTPYVDNLAAEILPAELRETLRLQLENKR	470
Db	185	RDSLSREANEELRCAQLQPGRLGTLQADPSLDPTSGLENLAAEILPAELRETLRLQLENKR	244
QY	471	LCRQEAADRERQBELORHLEDANRARGHLETHRLNQOQLSELRAQVEDLOKALOEGGK	530
Db	245	LCQGEAADREERQBELORHLEEDANRARGHLEAQRQLNQOQLSELRAQVEBELKALOEGGK	304
QY	531	TEDASITLLKRKLEEHLOKLHEADLELQKRREYIELEPPTDSSTARRIEELQHNLOKKD	590
Db	305	TED-PTLLKRKLEHLOKLHEADLELQKRREYIELEPPTDSSTARRIEELQDSLOKKD	362
QY	591	ADLRAMEERYRRYVDKARVMYQMTPMKRPAAGAPPELHSLRQLRERDVRIRHLEMDFE	650
Db	363	ADLRAMEERYRRYVDKARTVIQLEPKRPPTVVSPEFHTLASQLWERNLRIRQEMDYE	422
QY	651	KSRSRQEQEKLISAWYNGMALQORACEERAPAHQAQSFQAQRRLATNSRGRPLGRLAS	710
Db	423	KSRRQEQEKLISAWYSGMALEHRAGEHAPAHQAQSFQAQRRLATNWRGRPLGRQA	481
QY	711	LNLRPTDKH 719	
Db	482	LSLRPTDKH 490	
RESULT 5			
Q8R347 PRELIMINARY; PRT; 473 AA.			
AC	Q8R347		
CD	01-JUN-2002 (TREMBLrel: 21, Created)		
DT	01-JUN-2002 (TREMBLrel: 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel: 21, Last annotation update)		
DE	Similar to hook2 protein (Fragment).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid:10090;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RA	Strausberg R.J.		
RL	Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases..		
DR	EMBL; BC026609; AAH26609.1;		
FT	NON_TER		
SQ	SEQUENCE 473 AA; 55950 MW; C035938BD9B106D5 CRC64;		

Query Match	56.3%;	Score	2048.5;	DB 11;	Length	473;			
Best Local Similarity	86.9%;	Pred. No.	2.5e-89;						
Matches	413;	Conservative	28;	Mismatches	32;	Indels	3;	Gaps	
Qy	244	OLBOLQENFRLESGDEDFRLRCALEREVAELQHNQALTSLSAQAKALKQEMDELR	303						
Db	1	OLBOLQENFRLESSREDDRLCLELEREVAELQQNQALTSLSAQAKALKQEMDELR	600						
Qy	304	SERAGOLEATLTSCRRRLGELRLRQVROLEB RNAGHAERTROLDELRAGSLRAQLE	363						
Db	61	SERARQLETTLNSCRRRLGSELRLRQVROLEB RNAGHAERTQLSEELRRAGSLRAQLE	120						
Qy	364	AQRQVOVLOGQOQEEAMKAEMKLFECRNLBEEKYVESVTEKERLLAERDSLREANEELRC	423						
Db	121	AQRQVOVLOGQOQEEAMKAEMKLFECRNLBEEKCDLVTEKERLLTTERUSLREANEELRC	180						
Qy	424	AQOLPRLGTQADPSLDPTSTPVDNLAAILPAELRETLLRLQLENKRLCRQEAANDREOE	483						
Db	181	AQOLPRLGAQADLSLDPTPSGLENLAAAILPAELRETLLRLQLENKRLCQOEAAANDREOE	240						
Qy	484	ELQRHLEADNARHGLETQHRLNQOQLSELRAQVEDLQKALQEQGKTDATISILKRLK	543						
Db	241	ELQRHLEAANRARGHLEAQORLNNQOQLSELRAQVEELQKALQEQGKTD--PTLLKRLK	298						
Qy	544	EEHLQKLEHADLELQKRREYIEELPPTDSTARRTEELQHNLOKXDADLRAMEERYRY	603						
Db	299	EDHLQKLEHADLELQKRREYIEELPPTDSTARRTEELQDSLOKXDADLRAMEERYRY	358						

Qy	364	AQRQVOELQGOQEEAMKA	EKWLFCERNLEEKY	ESVTKERELLAE	RDSLREANEELRC	423	
Db	121	AQRQVOELQGOQEEAMKA	EKWLFCERNLEEKCDL	VTKEKRLT	RLSRREANEELRC	180	
Qy	424	AQLOPGLTOADPSLSD	TPVDNLA	EILPAELRET	LLRLQLENKRLCQEA	ADRRQE 483	
Db	181	AQLOPGLAQADSLSD	PTPSGLE	NLA	EILPAELRET	VLRLQLENKRLCQEEA	ADRRQE 240
Qy	484	ELQRHLEDANR	RHGLETOHRLNQOOL	SELRAQVEDI	LQKALQOGGKT	EDATSI	LLKRL 543
Db	241	ELQRHLEBANR	RHGLEAQORLNQOOL	SELRAQVEEL	LQKALQOGGKT	ED--PT	LLKRL 298
Qy	544	EEHLQKLEADLE	LQKRREYIEELP	PTDS	TARRIEELQHNLQK	DADLRA	MEERYRY 603
Db	299	EDHLQKLEADLE	LQKRREYIEELP	PTDS	TARRIEELQDS	LQKDALRA	MEERYRY 358

QY 604 VDKARMVOTMHPKORPAAGAPPELHSLRTQLRDVRIRHLEMDFKSRSQREBEKLL 663
 DB 359 VDKARTVOTLEPRPPTVVSFHTLSQWERNLRIQEMENDYKRRRQREBEKLL 418
 QY 664 ISAWNMGMAQOORAGRAPAHQAQSLAORLATNSRRGPGPLASINLRPTDKH 719
 DB 419 ISAWNMGMALEHRAHEHAPAHQAQSLAORLATNARRGPGPLRQA-LSLRPTDKH 473

RESULT 6
 Q9BY13 ID Q9BY13 PRELIMINARY; PRT; 718 AA.
 AC Q9BY13
 DT 01-JUN-2001 (TremBrel. 17, Created)
 DT 01-JUN-2001 (TremBrel. 17, Last sequence update)
 DT 01-JUN-2002 (TremBrel. 21, Last annotation update)
 DE Golgi-associated microtubule-binding protein HOOK3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21135884; PubMed=11238449;
 RA Walenta J.H., Didier A.J., Liu X., Kramer H.;
 RT "The Golgi-associated hook3 protein is a member of a novel family of
 RT microtubule-binding proteins.";
 RL J. Cell Biol. 152:923-934 (2001).
 DR EMBL: AF241830; AAK29204.1;
 DR PRINTS; PR01608; BACINVASINC.
 SO SEQUENCE 718 AA; 83125 MW; 9528BC9C854D39FA CRC64;

Query Match 48.3%; Score 1758; DB 4; Length 718;
 Best Local Similarity 49.9%; Pred. No. 1.8e-75;
 Matches 360; Conservative 158; Mismatches 107; Indels 16; Gaps 8;

QY 2 SVDAELGCSLLTWTQTHVSPCASPDLSGLAVYLVNQIDPSWFWNEAWLOGISED 61
 DB 6 SLERAELCESLLTWTQTHVSPCASPDLSGLAVYLVNQIDPSWFWNEAWLOGISED 65
 QY 62 GPNKLVNKLKWLRSVAVESQDVLAHPVSEHLDPVSLIGFSDPAELCKLLQLVLGC 121
 DB 66 GDNRLKISNKKKILGIDTINHEILGQIINDFTLPVNLIGHSDDAELGRMLQLILGC 125
 QY 122 ATSCCKQDHIQRIWLTESVQHVYVMEAIQELMTKDTSPDLSPTGYNFDQSRRYFLS 181
 DB 126 AVNCEQKEVYQIAIMMEESVQHVYVMTAIELMSKSPVAGNDAYVDLDQLKK--TT 182
 QY 182 EEAEG-----DELQORCLDLRLQLMLLSEKQSLAQENAGLRMRGREGTGLTA-K 236
 DB 183 BELNEALSAKEEIAQORCHELDNQVAAQEEKSSLLAENQVLMERLNQSDSIEDPNSAGR 242
 QY 237 KLLLLQSOLEQLOEFTLESGREDELRCAELEREVEALQHRNQAULTSLAQEAQAKDE 296
 DB 243 RHLQLOTQLEQLETFLEAKQDYRICELEKISELQOQDELTTLDAEQAQSLADE 302
 QY 297 MDLROSSERAGOLEATUTSCRRRLGELRLRROVQLEERNAGHAETROLEDELRAG 356
 DB 303 IDVLHRSDDKSVKLEGGVYSKKLEDGLDRLRQVKKLEENQVMTQVTSLEELRAN 362
 QY 357 SLRAQLERQROVQLOQROEAMKAELKWLFECEKNLEEKYESTVKEKERLLAERDLSRE 416
 DB 363 AARSLEYTKROVQLOVQLOVQLOVQLOVQLOVQLOVQLOVQLOVQLOVQLOVQLO 422
 QY 417 ANEELRCAQLQPGTQADPSLOPTSTVFNLAABILPAELRETLRLRLOLENKRL-CROE 475
 DB 423 TTEELRCVQAQEGQLTQGLPLGSOESSDLSAAEIVTPETREKLRILQHENKMLKNOE 482
 QY 476 ADREERQELQRLHEDANRAGHLETOHRLNQOOLSELRAOVEDLOKALOEGGKTEAI 535
 DB 483 GSDNEKIALQLSLDDANLRKNELETENRLVNRQLEVOQVEELQKLSQDOOGSKAED-- 540

QY 536 SILKRRLEHKLQHEADLELQKREYIEBELPPTDSSTARRIEELQHLNKKQADLRA 595
 DB 541 SVLLKKLEHKLQHEADLELQKREYIEBELPPTDSSTARRIEELQHLNKKQADLRA 599
 QY 596 MEERYRYVDKARMVOTMHPKORPAAGAPPELHSLRTQLRDVRIRHLEMDFKSRSQ 655
 DB 600 MEERYRYVDKARMVOTMHPKORPAAGAPPELHSLRTQLRDVRIRHLEMDFKSRSQ 657
 QY 656 REBEKLLISAWNMGMAQOORAGRAPAHQAQSLAORLATNSRRGPGPLASINLR 713
 DB 658 REMEKYIVSAWYVNMGMTHLHKAEDRLASTGSGQSLARQATQATQATQATQATQAT 717

QY 714 R 714
 DB 718 R 718

RESULT 7
 O60561 ID O60561 PRELIMINARY; PRT; 728 AA.
 AC O60561; Q9UCJ3.
 DT 01-AUG-1998 (TremBrel. 07, Created)
 DT 01-AUG-1998 (TremBrel. 07, Last sequence update)
 DT 01-JUN-2002 (TremBrel. 21, Last annotation update)
 DE Hook1 protein (DJ782L23.1).
 GN HOOK1 OR DJ782L23.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99126435; PubMed=9927460;
 RA Kramer H., Phistry M.;
 RT "Genetic analysis of hook, a gene required for endocytic trafficking
 RT in Drosophila.";
 RL Genetics 151:675-684 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-581 FROM N.A.
 RA Wallis J.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF044923; AAC09298.1;
 DR EMBL; BC011621; AAHL1621.1;
 DR EMBL; AL035416; CAB52263.1;
 DR InterPro; IPR002017; Spectrin.
 SO SEQUENCE 728 AA; 84647 MW; E35CA91F2BB89B3E CRC64;

Query Match 45.8%; Score 1667; DB 4; Length 728;
 Best Local Similarity 48.6%; Pred. No. 3.6e-71;
 Matches 347; Conservative 141; Mismatches 202; Indels 24; Gaps 8;

QY 5 KAEICGSLTWTQTHVSPCASPDLSGLAVYLVNQIDPSWFWNEAWLOGISED 64
 DB 11 KLPLCLSLMIWLTQTHVSPCASPDLSGLAVYLVNQIDPSWFWNEAWLOGISED 70
 QY 65 WKLKLVNKLKWLRSVAVESQDVLAHPVSEHLDPVSLIGFSDPAELCKLLQLVLGC 124
 DB 71 WRKASNVKVLQGISVYHEFLGQOISEALIPDLNQITECSDPVVELGRLLQLILGC 130
 QY 125 CEKQDHIQRIWLTESVQHVYVMEAIQELMTKDTSPDLSPTGYNFDQSRR-YVFLSEE 183
 DB 131 CEKQDHIQRIWLTESVQHVYVMTAIELMSKSPVAGNDAYVDLDQLKK--TT 190
 QY 184 ABEGLQORCLDLRLQLMLLSEKQSLAQENAGLRMRGREGTGLTA-K 241
 DB 191 LAEKELQORCLDLRLQLMLLSEKQSLAQENAGLRMRGREGTGLTA-K 250
 QY 242 OSLEQLQOENFRLESGREDELRCAELEREVEALQHRNQAULTSLAQEAQAKDE 301

```

Db 251 QLOLEQLEENFRLEAAKDDVVRHCELEKQLEIFQHRNDELTSLAETRAKLDIDVLR 310
QY 302 OSSERAGOLEATLTSRRRLGELRELAROVQLEERAGHAERPROLEDELRRAGSLRAQ 361
Db 311 ATSDKANKLESTVEIYKQKQDLNDLRKQVKTLOETNMVMHNTVSLSEELKANAARTQ 370
QY 362 LEAORROVQLEQOROEAKAEKWLFEKNLEKESVTKERKLLAERDSLEANEEL 421
Db 371 LEYKROVQDLHVKLSESKRAOTLAEMKRELEKHEALKKEKRLTEORTLTETNEEL 430
QY 422 RQAOLQPRGTQADPSLDPTSTPDVNLAAETLPAELRETLLRLQLENKRL-CRQEAADRE 480
Db 431 RCOVQOQDLHNTQDAS---ATKSYLENLAETMPVEYFIRLQENKMLRLQOEGSENE 487
QY 481 ROELQRLHEDANRARGLETOHRLNQOOLSELRAQVEDLQALQOEGGTEDAISILK 540
Db 488 RIBELQOLQKHKRMKLETELQRLSERINLEQOQLELQSLQOEGSKSESSSKLK 547
QY 541 RKLHEHLQKLEHLELQKREYTELEPPTDSSTARIELELQHLNLOKQADLRAMEERY 600
Db 548 OKLEAHMEKLETEVHEELQKQELIEDLPDINQNVKINELEALQKQKDEDMAMEERY 606
QY 601 RRYVDKARVMQTMPEKQORPAAGAPPELHSLRTOLEDRVRIRHLEWDFEKSORSOE 660
Db 607 KMYLEKARNVKTLDPLNPASA---EIMLLKQLEAKERRIETLESECKVAKF-RDYEE 662
QY 661 KLLISANYNGMALQORAGEER-----APHAQSFLAQORLATNSRR 702
Db 663 KLIVSAWYNKSLAQKLGMSRLVGGGACSDTGACTPARSFLAQORHITNRR 716

RESULT 8
061493
ID 061493 PRELIMINARY; PRT; 678 AA.
AC 061493;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Hook protein.
GN HOOK OR HK.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RA Phistry M., Sunio A., Kramer H.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF044926; AAC09301.1;
DR FlyBase; FBgn0024376; Dvir\hook.
SQ SEQUENCE 678 AA; 77198 MW; 1B8535B80F06C673 CRC64;

Query Match 23.7%; Score 864; DB 5; Length 678;
Best Local Similarity 31.6%; Pred. No. 2.3e-33;
Matches 230; Conservative 155; Mismatches 255; Indels 88; Gaps 19;

QY 11 SLTLWLTFFVSPCASQDSSGLAVYVNLQIDPSWFNEAWLQGI-SEDPGNWKLKV 69
Db 10 SLLEWFTLNLNAPHANAEADGVALAQAALQAPESFTNSLKSITSSAVGNGWRLRM 69
QY 70 SLNKWVRLSVESQDVLAPVSEHL-PDVSLLIGESDPAELGKLQVLGCAISCEKK 128
Db 70 SLNKVYGVYEVYSYDLYNTLQHDVFKVPDQVIAECRDLSELERLLQLVLGCAVCAKK 129
QY 129 QDHTORTIMTEESVQHVVMIAQL--MTKDTDP-----SLSPETVGNFDSQSR 176
Db 130 QSYICEIMCELEQANIMRALQLESSTROTTEGGVVSSLSRNSLMDLGN----- 182
QY 177 YVFLSEAEDEQLQRCLOLERLMLLSEBKQSLAQENAGLRMRGPRGE--GTPLT 234
Db 183 ---AKALEERDAAKQCFETEKMKLLIDETKNLQELHKLQLEFARLEHTIGDGV 238

```

```

QY 235 AKLL-----LQSOLEQLEENFRLESGREDRLRCAELEREVAELQHRNQAITSLA 287
Db 239 LGPTQASVRYNEURQLVKEELQSEGAREDLKIAQOQETDLHMQORIDELMKST 298
QY 288 QEAQALDEMDQLQSSERAGOLEATLTSRRRLGELRELAROVQLEERAGHAERTRQ 347
Db 299 AELTALDEVDVLRSTDKLVCEAQLQETKYLEYNDLKKHKVLMLEERSADYVQNAQ 358
QY 348 LEDELRRAGSLRAQLEAORROVQLEQOROEAKAEKWLFEKNLEKESVTKERK 407
Db 359 FEEDAKRYANTKGOVFLFKKEIQDLHAQLDNESKNVLEKFNKLNESKTLALQREK 418
QY 408 LAERDSLEANEELRCAQLQPRGTQADPSLDPTSTPDVNLAAETLPAELRETLLRLQ 467
Db 419 LKERDNLRAEPDELKCGOL-----STNSGLTGTMTSRELQPPAMDMKQRLQEE 468
QY 468 NKRLCROEAADREQEELQRLHEDAN-RARHGLETOHRLNQOOLSELRAQVEDLQALQ 526
Db 469 NKALREGGG---QTLAQQLDDANKRCEHLRQLKANTANERILSLSHASQD----- 517
QY 527 QGGYEDAISILLKRLKEHLQKLEHLELQKREYTELEPPTDSSTAR-RTEELQHN 585
Db 518 -----DPTLKENEFKQIKQLMELN---EOKTLQIE--ESATQNSTMQCKITQLEST 564
QY 586 LQKQDADLRAMEERYRYVDKARVMQTMPEKQORPAAGAPPELHSLRTOLEDRVRIRH 645
Db 565 LATREQELMAVEYKIRKIERAKEVIRKIDPR---TASVMEANNL-----EKSDV--- 612
QY 646 EMDFEKRSRQREKEKLLISAWYNGMALQORAGEERAP---AHAQSFLAQORLATNSRR 702
Db 613 -IEESKTKSGMEEQLMASAFYRLGVNAQORDAVDSKLALLMGSGQTFLARQ--QSAPR 669
QY 703 GPLGLAS 710
Db 670 KPLTMTKS 677

RESULT 9
09VJ15
ID 09VJ15 PRELIMINARY; PRT; 679 AA.
AC 09VJ15;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE HK gene product (LD05265p).
GN HK OR CG10653.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

```


RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Mount D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reshe M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.,
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003661; AAF53742.1;
 DR EMBL: AF069337; AAL39482.1;
 DR FlyBase: FBgn0001202; hk.
 SQ SEQUENCE 679 AA; 76664 MW; COAB51E0AAD285AC CRC64;

Query Match 23.5%; Score 854; DB 5; Length 679;
 Best Local Similarity 31.7%; Pred. No. 6.7e-33;
 Matches 227; Conservative 156; Mismatches 261; Indels 72; Gaps 17;

QY 1 MSVDKALGSLTLTQTHVPSPCASPDLSGLAVAYVYVNLQIDPSWFNEAWLOGISED 60
 DB 1 MSAPKNEMYSLLEWFKTLNLNAPHADAESLADGVALAQAALNFAFESFTDMLSKIKAS 60

QY 61 P-GPNKLVKSNLKVLSVEYSQDVLAHPVSEHLPDVSLLGFEPSDPAELKLLQLVL 119
 DB 61 AVGSNNRLRMSNLKKVTSQVYDVSVNLNYSDFSQPDQLQIAEKCDLGELERLQLVL 120

QY 120 GCAISEKKODHTORIMTLSESVQHVYVMEAIQELMTKDTDPDSLSPETYNFDSQSR--RY 177
 DB 121 GCAVNCNAEKOSYITEIMCLSEELQANIMRALQEL--ENTROASTPEG-GVASSLSRGSRT 177

QY 178 YFLSEEA--EEGDELOQRCIDLERQLMLLSEKQSLAQENAGLRMRGRPEGEET----- 230
 DB 178 GLDLSKAVQEDRALAQKCFETEKMLLLIDEXTNLQOELHKLQOEFARLEQHSVTIGDD 237

QY 231 -----PGLTAK-KLLIQSOLEQLOEENFLESGREDELRCAELEREVAELQHRNQALT 284
 DB 238 GVSIGPVQTSVRYNELRRQDLDLKEELLQSEGARKDLKIKAQOQDLDLLHMQRIEELM 297

QY 285 SLAQEAQALKDEMDLROSSERAGQELATITSCRRRLGELRELRRVQROLEERNAGHAER 344
 DB 285 SLAQEAQALKDEMDLROSSERAGQELATITSCRRRLGELRELRRVQROLEERNAGHAER 344

QY 298 KSSAEVTTLDKDEVDLRESNDKLCICEAQDLYKKKLEDYNDLKKQVKKILEERSADYVQ 357
 DB 298 KSSAEVTTLDKDEVDLRESNDKLCICEAQDLYKKKLEDYNDLKKQVKKILEERSADYVQ 357

DB 467 EAENKALREGOGG-----QTALAQLDDANKRCENIREOLKATANERILSLSHASQSDPIL 522
 QY 524 LQEOGKGTEDAISILLKKLEHLOKLEADLEQKREYIEELPPDTSSTAR-RIBEL 582
 DB 523 KESEFGK-----IKQLMELNEK-----TLQLE-----EAVTOSTLQCKVTOL 562
 QY 583 QHNLOKQADADLAMEERYRYVYDKARMVMQTMPEKORPAAGAPPELHSLRTOLRERDRI 642
 DB 563 ETNLSAREQEILVDAKYKCKVEKAKVKSIDPRIASA-----LDASVLEKSA 613
 QY 643 RILEMDFEKSQRQOEKELLISANYNMGMALOQORAGEERAP---AHQSELAQOR 695
 DB 614 -----VEEPEPKMSYMEQALMTSAFYRLGVNAQORDAIDSKLAILMGSGQTFLARQ 665

RESULT 10
 Q24185 PRELIMINARY; PRT; 679 AA.
 AC Q24185;
 DT 01-NOV-1996 (TREMELrel. 01, Created)
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
 DE HOOK.
 GN HK OR HOOK OR CG10653.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96281964; PubMed=8682859;
 RA Kramer H., Phistroy M.;
 RT "Mutations in the Drosophila hook gene inhibit endocytosis of the boss transmembrane ligand into multivesicular bodies.";
 RJ J. Cell Biol. 133:1205-1215(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Phistroy M., Sunio A., Kramer H.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U48362; AAC47261.1;
 DR EMBL: AF044925; AAC09300.1;
 DR FlyBase: FBgn0001202; hk.
 SQ SEQUENCE 679 AA; 76663 MW; 8415964DF7CBB875 CRC64;

Query Match 23.4%; Score 850; DB 5; Length 679;
 Best Local Similarity 31.6%; Pred. No. 1e-32;
 Matches 226; Conservative 157; Mismatches 261; Indels 72; Gaps 17;

QY 1 MSVDKALGSLTLTQTHVPSPCASPDLSGLAVAYVYVNLQIDPSWFNEAWLOGISED 60
 DB 1 MSAPKNEMYSLLEWFKTLNLNAPHADAESLADGVALAQAALNFAFESFTDMLSKIKAS 60

QY 61 P-GPNKLVKSNLKVLSVEYSQDVLAHPVSEHLPDVSLLGFEPSDPAELKLLQLVL 119
 DB 61 AVGSNNRLRMSNLKKVTSQVYDVSVNLNYSDFSQPDQLQIAEKCDLGELERLQLVL 120

QY 120 GCAISEKKODHTORIMTLSESVQHVYVMEAIQELMTKDTDPDSLSPETYNFDSQSR--RY 177
 DB 121 GCAVNCNAEKOSYITEIMCLSEELQANIMRALQEL--ENTROASTPEG-GVASSLSRGSRT 177

QY 178 YFLSEEA--EEGDELOQRCIDLERQLMLLSEKQSLAQENAGLRMRGRPEGEET----- 230
 DB 178 GLDLSKAVQEDRALAQKCFETEKMLLLIDEXTNLQOELHKLQOEFARLEQHSVTIGDD 237

QY 231 -----PGLTAK-KLLIQSOLEQLOEENFLESGREDELRCAELEREVAELQHRNQALT 284
 DB 238 GVSIGPVQTSVRYNELRRQDLDLKEELLQSEGARKDLKIKAQOQDLDLLHMQRIEELM 297

QY 285 SLAQEAQALKDEMDLROSSERAGQELATITSCRRRLGELRELRRVQROLEERNAGHAER 344
 DB 298 KSSAEVTTLDKDEVDLRESNDKLCICEAQDLYKKKLEDYNDLKKQVKKILEERSADYVQ 357

Qy	345	TRQLEDELRAGSLURAOLENAQRQVQELQQRQOEAMKEWLFECRNLEEKYESVTKK	404
Db	358	NAOPFEADKYANTKGOVELFKKEIQDLHAKLDAESSKNVKEFDKNKLESNIALQRAK	417
Qy	405	ERLLAERDSLRANEELRCQAOLPGRGLQTQADPSLDSTSPVDNLAELIIPAEURETLLK	464
Db	418	DSLKERDNRLEAVDELKQCQLSSN-----TALTGTTVSRELOPSATVEKQLR	466
Qy	465	QLENKRCROEAAADREORELOQLHLEDANRAHGLETOHR-LNQOQLSELRAQVEDLQRA	523
Db	467	EAENKALREQGG-----QTALAQLODDANKRCENREQLKTNERILTSLSHASQSDPIL	522
Qy	524	LOEQGGTEDAISLLKRLKEELHQKLHEADLELQKREVIEELEPPTDSSTAR-RIEL	582
Db	523	KESEFGKQ-----IKQLMELNEQ-----TLQLE-----EAVTQSTSLQCKVTOL	562
Qy	583	QHNLOKQKDALRAMEERYRYVDKARMVOTMEPKORPAAGAPPELSURTOLRERDVRI	642
Db	563	ETNLSAREQBIYVDKYRCVKEKAEVTKSIDPRIASA-----LDASVLEKSADL	613
Qy	643	RHLEMDPEKRSRQREOEKLLISAWYNGMANQOORAGEERAP---AHAQSFLAQOR	695
Db	614	-----VEEPPKPKSVMEQWMTSAFYLGVNQAORDAIDSKLILMGSGOTTFLAAR	665

```

RESULT 11
QC9CTNG
ID          PRELIMINARY:      PRT: 256 AA.
AC          QC9CTNG;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE A930033L17RIk protein (Fragment).
GS A930033L17RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=RETINA;
RC MEDLINE=21085660; PubMed=112117851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Anono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann C., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK020924; BAB32257.1; "-
DR MGD; MGI:1925213; A930033L17RIk.
FT NON_TER 256
SQ SEQUENCE 256 AA; 29278 MW; 3AB6C1F2E1F85354 CRC64;

```

Query Match 16.3%; **Score** 593.5; **DB 11;** **Length** 256;
Best Local Similarity 48.6%; **Pred. No.** 4.5e-21;
Matches 118; **Conservative** 46; **Mismatches** 76; **Gaps** 3; **Gaps** 2

QY . 8 LCGSLLTWLTFHFVPCASPDQLSSGLAVAYIVLNQDPSWFNEAWLOGISEDPGGNNKL' 67

```

|| ||| |||||| || : ||: ||| ||| | ||| ||| : ||| ||| :

```

[illegible]

```

QY 381 MKAQKWLFECENLKESVTKERLLAERDSL-REANE-----ELRCAQLQPRGLTQAD 435
Db 342 MAQKQSDHSEHLGWELEQISRTSELSEAPKSLGHEVNELTSSRLKLENNOSLTKTV 401
QY 436 PSLDPTSPVDNLAELPAELRETLRLQLENKRL-----CRQE 475
Db 402 EELRTTVDSVEGNASKI-----LKMENQRLSKKVEILENEIYQKSLQNCNL 452
QY 476 AAD-----RERQELQRLHLEDANRARGHLETOHRLNQOOLSELRAQVEDL 520
Db 453 SKDLMKERAKAQLKTIETURENSERQIKILEQEN--EHLNQTVSSLRQSRQISAEARVKDI 510
QY 521 Q---KALQEOGGKTEDAISL-----LKRLEELHKLHLEADLELQKREYIELEPPT 571
Db 511 EKENILHESIKETSSKSLKIEFEKROIKKLEHYKEGERAE-ELENELHLEKENELL 569
QY 572 DSS-----TARRIBELQNLQKDDADLAMEERYRYVDKARMVMOTMEPKQRPAAQAP 625
Db 570 QKKITNLKITEKIEA-----LEQENSELERENRKLKKTLDGFKNTFQLESLEKENSQID 625
QY 626 PELHSLR---TOLRERDVRIRHLEMDFEKRSQRQEEK 661
Db 626 EENLELRNVSCLKASMKMAQLOLENKELESEKQLKK 664

RESULT 13
Q9UI01 PRELIMINARY; PRT; 742 AA.
AC Q9UI01
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Hypothetical 87.3 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HYPOPHALAMUS;
RA Shi J., Jiang C., Huang C., Ren S., Li Y., Zhou J., Yu Y., Xu S.,
RA Wang Y., Fu G., Chen Z., Han Z.;
RT "A novel gene expressed in human hypothalamus.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF112218; AAF17206.1;
DR HSSP; P03069; IGCL.
KW PRINTS; PRO1608; BACINVASING.
KW Hypothetical protein.
SQ SEQUENCE 742 AA; 87326 MW; F54FC3E5B2D0B4E2 CRC64;

Query Match 10.6%; Score 385.5; DB 4; Length 742;
Best Local Similarity 23.6%; Pred. No. 8.6e-11;
Matches 165; Conservative 141; Mismatches 248; Indels 145; Gaps 26;

QY 71 NLKMWRLSVESQDVLAHPVSEHLPDVSLSICE--FSDPA--ELGKLQVLGCAISCE 126
Db 3 NLSILVROIKFYQETLQQLIMMS-LPNVLIIGNPFSEQGTVEVKRLLLLLGCAVQCQ 61
QY 127 KKQDHQIRIMTLESVQHVYMEAIQELMTKTDPSLSPEYGN--FDSQSRRYFYLSSE 183
Db 62 KKEEFIERIOGLDFDKAAVAHHIOEV-----THNENVFQDQWMEVTDMSQE 109
QY 184 AEEDEQLQRCLOLEROLMLLSEKQS---LAQENAGL-----RERMGRPEGETPGL--- 233
Db 110 DIE-PLKKNALHLKRLIDERDEHSEFIELSEERDGLHFLPHASSAQSPGSGPKMT 168
QY 234 -----TAKKULLQSLQEQLOEQENFRLSREDERLRCABLEREVAELRHQ 281
Db 169 ESROHLSVELADAKAKIRRLRQLEETE-Q--LLDCKOELE-----QMEIELKRLQENM 221
QY 282 ALTSIAQEAQAKDMEDELROSSERAGOLATLSCRRRLGELRELRRQVRLQLEERNAGH 341

```

```

Db 222 NLLSDARSARMYRDDELALREKAVRVKLESEVSRYKERLHDIEFYKARVEELKEDNOVL 281
QY 342 AERTRQLEDEL--RAGS-----LRAQL-----EAQRROVQLOQRQEEA 380
Db 282 LETKTMLEDOLEGTRARSQKLEHEKENLQKALHDMEMERDMRKKIEELMEENMTLE 341
QY 381 MKAQKWLFECENLKESVTKERLLAERDSL-REANE-----ELRCAQLQPRGLTQAD 435
Db 342 MAQKQSDHSEHLGWELEQISRTSELSEAPKSLGHEVNELTSSRLKLENNOSLTKTV 401
QY 436 PSLDPTSPVDNLAELPAELRETLRLQLENKRL-----LKMENQRLSKKVEILENEIYQKSLQNCNL 452
Db 402 EELRTTVDSVEGNASKI-----LKMENQRLSKKVEILENEIYQKSLQNCNL 452
QY 476 AAD-----RERQELQRLHLEDANRARGHLETOHRLNQOOLSELRAQVEDL 520
Db 453 SKDLMKERAKAQLKTIETURENSERQIKILEQEN--EHLNQTVSSLRQSRQISAEARVKDI 510
QY 521 Q---KALQEOGGKTEDAISL-----LKRLEELHKLHLEADLELQKREYIELEPPT 571
Db 511 EKENILHESIKETSSKSLKIEFEKROIKKLEHYKEGERAE-ELENELHLEKENELL 569
QY 572 DSS-----TARRIBELQNLQKDDADLAMEERYRYVDKARMVMOTMEPKQRPAAQAP 625
Db 570 QKKITNLKITEKIEA-----LEQENSELERENRKLKKTLDGFKNTFQLESLEKENSQID 625
QY 626 PELHSLR---TOLRERDVRIRHLEMDFEKRSQRQEEK 661
Db 626 EENLELRNVSCLKASMKMAQLOLENKELESEKQLKK 664

RESULT 14
Q8SX64 PRELIMINARY; PRT; 1235 AA.
AC Q8SX64
DT 01-JUN-2002 (TREMELrel. 21, Created)
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE LD40094p.
GN CG12734.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY094827; AAM1180.1;
SQ SEQUENCE 1235 AA; 142401 MW; C7C813ECBD86C58A CRC64;

Query Match 10.4%; Score 378; DB 5; Length 1235;
Best Local Similarity 22.3%; Pred. No. 3.3e-10;
Matches 178; Conservative 145; Mismatches 321; Indels 154; Gaps 31;

QY 1 MSYDKAELCGSLTWTQTFHVPSPCASPODLSGLAVAYVLNOIDPSWFNEAWLQIGISE 60
Db 10 MEIDEP-VNFSIIISWLES-----CLPRAELLAGYT-----SLLDGHIHSVWLQ---ID 54
QY 61 PGP-----NWK-----LKVSNLKMWRLSVESQDVLAHPVSEHLPDVSLSICEFS 106
Db 55 PEQONNPSELNDLNGKSLSTARAKNFECVVRNLKSFPEELGOTILV--LPDAFTLGHHP 112
QY 107 DP-----AELGKLQVLGCAISCEKKQDHQIRIMTLESVQHVYMEAIQEQ-----LMT 155
Db 113 ESKNGLEQNKTLTLLGAAVQCPNKELFIARIKELDLETOHAIVGLIKQVTDHSLVLT 172

```


Db 427 TMDLKSFSENEEDCNSGDSISEQLTNNAQTRALKLELENRLTALEQLKESFHEST 486
Qy 486 QHLE-DANRAHGLE-TOHRLNQOOLSELRAQVEDLOKALQEOGGKTEDAISILLKRL 543
Db 487 SKMLEKEKKKLSLKIEQMOENINRLTQQNVELEGVFNALKEENKKLQDAVDNRQKSYD 546
Qy 544 BEHLOKLEADLE-LORKREYIEELEPTDSSTARIEELOHNLQKKDADLRAMEE---- 598
Db 547 ROSLER--EADQKLSDAEQHVETL-----NKEKQRIOTLNEIORRADDLERLAESKTK 599
Qy 599 RYRRYVDKARVMQOT-----MEPKQPPAAGAPPELHSLRTQLRERDVR-----IRH 644
Db 600 ELEQLKSRQVELTKQKLYEIEARVSTYERENASLLKEVSKLKEGSEKSVOLDSDSINR 659
Qy 645 LEMDFE-----KRSRSEQ-EEKLLISAWYNGMALQORAGEERAPAHQSFQAQRILA 697
Db 660 LDVQSKELQKLGKALEDSGEQVHOKLVELEKQNELA-SQRIIDQEMISTLRNDLVTGTIV 718
Qy 698 TNSRRGPIGLRLASLNRP 715
Db 719 TKKVRNLEKGLADEEP 736

Search completed: March 4, 2003, 15:02:39
Job time: 53.2588 secs

GenCore version 5.1.1.3
Copyright. (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:51:02 ; Search time 12.4234 Seconds
(without alignments)
1732.709 Million cell updates/sec

Title: US-09-763-902B-3

Perfect score: 2620

Sequence: 1 MAAREPPLGDKPTDFED.....TQQLIKYWEALPEAKAIA 519

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2612	99.7	519	SNX2_HUMAN	O60749 homo sapien
2	2564	97.9	519	SNX2_MOUSE	Q9CWX8 mus musculu
3	1564.5	59.7	522	SNX1_RAT	Q9N27 rattus norv
4	1557.5	59.4	522	SNX1_MOUSE	Q9WV80 mus musculu
5	1534.5	59.3	522	SNX1_HUMAN	Q13596 homo sapien
6	386	14.7	675	VPS5_YEAST	Q92331 saccharomyc
7	308.5	11.8	404	SNX5_HUMAN	Q9Y5X2 homo sapien
8	305.5	11.7	404	SNX5_MOUSE	Q9D8U8 mus musculu
9	300.5	11.5	387	SNX7_HUMAN	Q9UNH6 homo sapien
10	291.5	11.1	387	SNX7_MOUSE	Q9CY18 mus musculu
11	248	9.5	406	SNX6_HUMAN	Q9UNH7 homo sapien
12	228	8.7	423	SNX4_YEAST	P47057 saccharomyc
13	221	8.4	450	SNX4_HUMAN	Q95219 homo sapien
14	214	8.2	465	SNX8_HUMAN	Q9Y5X2 homo sapien
15	200	7.6	595	SNX9_HUMAN	Q9Y5X1 homo sapien
16	179	6.8	162	SNX3_HUMAN	O60493 homo sapien
17	178	6.8	162	SNX3_MOUSE	O70492 mus musculu
18	169.5	6.5	162	SNX3_HUMAN	Q9UNY4 homo sapien
19	164.5	6.3	511	MYP1_YEAST	P40959 saccharomyc
20	164.5	6.3	2663	CENE_HUMAN	Q02224 homo sapien
21	163.5	6.2	893	YIN92_CAEEL	P34531 caenorhabdi
22	163	6.2	165	SNXC_MOUSE	O70493 mus musculu
23	162.5	6.2	992	SNXJ_HUMAN	Q92543 homo sapien
24	159.5	6.1	1208	PCP1_SCHPO	Q92351 schizosacch
25	158	6.0	944	NUF1_YEAST	P32380 saccharomyc
26	158	6.0	1035	TAC2_MOUSE	Q91JG0 mus musculu
27	156	6.0	1026	TAC2_HUMAN	O95359 homo sapien
28	155.5	5.9	270	SNXB_HUMAN	Q9Y5W9 homo sapien
29	154.5	5.9	591	YN48_YEAST	P42846 saccharomyc
30	154.5	5.9	1875	MLP1_YEAST	Q02455 saccharomyc
31	153	5.8	628	SNX1_HUMAN	Q961F0 homo sapien
32	152	5.8	1939	MYH6_HUMAN	P13533 homo sapien
33	150.5	5.7	961	VDP_BOVIN	P41541 bos taurus

RESULT 1

SNX2_HUMAN STANDARD; PRT; 519 AA.
AC O60749; O43650; P82862; Q9BTS8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 2.
GN SNX2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Kurten R.C., Leychikis Y., Wiley H.S., Gill G.N.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99038232; PubMed=9819414;
RT Haft C.R., de la Luz Sierra M., Barr V.A., Haft D.H., Taylor S.I.;
RT "Identification of a family of sorting nexin molecules and
RT characterization of their association with receptors.";
RL Mol. Cell. Biol. 18:7278-7287(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strauberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May be involved in several stages of intracellular
CC trafficking
CC -!- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF065482; AAC17181.1;
EMBL; AF043453; AAB59852.1;
EMBL; BC003382; AAH03382.1;
Gene; HGNC:11173; SNX2.
MIN; 605929;
InterPro: IPR001683; PX.
InterPro: IPR005329; Sorting_nexin.
Pfam: PF00787; PX; 1.
Pfam: PF03700; Sorting_nexin; 1.
SMART: SM00312; PX; 1.
PROSITE: PS00195; PX; 1.
Transport; Protein transport.
DOMAIN 140 269 PX.

ALIGNMENTS

34 150 5.7 1224 1 DYNA_CHICK P35458 gallus gall
35 147.5 5.6 1189 1 SCII_CHICK Q90988 gallus gall
36 147 5.6 917 1 YGJ3_YEAST P33148 saccharomyc
37 147 966 1 STI0_MOUSE O5098 mus musculu
38 146.5 5.6 1935 1 MYSS_CYPCA Q90339 cyprinus ca
39 146 5.6 1938 1 MYS_AEQIR P24733 aequipecten
40 146 5.6 3660 1 DMD_CHICK P11533 gallus gall
41 145.5 5.6 1966 1 MYSB_CAEEL P02566 caenorhabdi
42 143.5 5.5 1938 1 MYH6_RAT P02563 rattus norv
43 143 5.5 1962 1 MYSA_DROME P05661 drosophila
44 142 5.4 1102 1 MYSC_CHICK P29616 gallus gall
45 142 5.4 1283 1 OSH2_YEAST Q12451 saccharomyc

```

FT CONFLICT 331      F -> V (IN REF. 3).
FT CONFLICT 384      S -> A (IN REF. 3).
SQ SEQUENCE 519 AA: 58535 MW: 897DA7E15935C3A0 CRC64;

Query Match
Best Local Similarity 99.6%; Score 2612; DB 1; Length 519;
Matches 517; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAAREPPPLGDKPTDFELEDGEDLFTSTVSTLESSPSPASPAPLPAEDISANSNGPK 60
DB 1 MAAREPPPLGDKPTDFELEDGEDLFTSTVSTLESSPSPASPAPLPAEDISANSNGPK 60
QY 61 PTEVVLDDDDREDFAEATEEVSLSDSPEPILSESPSPAVPTPTTLIAPRIESKMSA 120
DB 61 PTEVVLDDDDREDFAEATEEVSLSDSPEPILSESPSPAVPTPTTLIAPRIESKMSA 120
QY 121 PVIFDRSREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSV 180
DB 121 PVIFDRSREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSV 180
QY 181 KRPFSDPLGHSKLASKYLHVGIYVPPAPEKSIIVGTMKVKGEDSSSTFEVEKRRRALE 240
DB 181 KRPFSDPLGHSKLASKYLHVGIYVPPAPEKSIIVGTMKVKGEDSSSTFEVEKRRRALE 240
QY 241 RYLQRTVKHPTLLQDPLDRLQFLESSELPRAVNTQALSGAGILRMVKNKAADAVNKMKTIMN 300
DB 241 RYLQRTVKHPTLLQDPLDRLQFLESSELPRAVNTQALSGAGILRMVKNKAADAVNKMKTIMN 300
QY 301 EDAWFEKQOQFENLQDQRLKHLVSVYALVCHRKELSSANTAAFAKSAAMLGNSDHTAL 360
DB 301 EDAWFEKQOQFENLQDQRLKHLVSVYALVCHRKELSSANTAAFAKSAAMLGNSDHTAL 360
QY 361 SRALSOLAEVEEKIDQLHQOAFADFYMFSELLSDYRLTLAAVKGVDHRMKCKQKVEDA 420
DB 361 SRALSOLAEVEEKIDQLHQOAFADFYMFSELLSDYRLTLAAVKGVDHRMKCKQKVEDA 420
QY 421 QITLLKKREAEAKMVANKPKDQKQAKNEIREWEAKVQOQERPEQISKTIRKEVGRFEK 480
DB 421 QITLLKKREAEAKMVANKPKDQKQAKNEIREWEAKVQOQERPEQISKTIRKEVGRFEK 480

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: May be involved in several stages of intracellular
CC trafficking.
CC -!- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AK002692; BAB22287.1; -
CC DR ENBL; AK005470; BAB24060.1; -
CC DR ENBL; AK010572; BAB27035.1; -
CC DR MGI; 1915054; Snx2.
CC DR InterPro: IPR001683; PX.
CC DR InterPro: IPR005329; Sorting_nexin.
CC DR Pfam; PF00787; PX; 1.
CC DR SMART; SM00312; PX; 1.
CC DR SMART; SM00195; PX; 1.
CC DR PROSITE; PS0195; PX; 1.
CC KW transport; Protein transport. PX.
CC FT DOMAIN 140 269
CC FT CONFLICT 245 245 R -> I (IN REF. 1; BAB27035).
CC FT CONFLICT 428 428 R -> P (IN REF. 1; BAB27035).
CC SQ SEQUENCE 519 AA: 58471 MW: 55DD0BB74E82CD82 CRC64;

Query Match 97.9%; Score 2564; DB 1; Length 519;
Best Local Similarity 98.1%; Pred. No. 3.8e-129;
Matches 509; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 MAAREPPPLGDKPTDFELEDGEDLFTSTVSTLESSPSPASPAPLPAEDISANSNGPK 60
DB 1 MAAREPPPLGDKPTDFELEDGEDLFTSTVSTLESSPSPASPAPLPAEDISANSNGPK 60
QY 61 PTEVVLDDDDREDFAEATEEVSLSDSPEPILSESPSPAVPTPTTLIAPRIESKMSA 120
DB 61 PTEVVLDDDDREDFAEATEEVSLSDSPEPILSESPSPAVPTPTTLIAPRIESKMSA 120
QY 121 PVIFDRSREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSV 180
DB 121 PVIFDRSREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSV 180
QY 181 KRPFSDPLGHSKLASKYLHVGIYVPPAPEKSIIVGTMKVKGEDSSSTFEVEKRRRALE 240
DB 181 KRPFSDPLGHSKLASKYLHVGIYVPPAPEKSIIVGTMKVKGEDSSSTFEVEKRRRALE 240
QY 241 RYLQRTVKHPTLLQDPLDRLQFLESSELPRAVNTQALSGAGILRMVKNKAADAVNKMKTIMN 300
DB 241 RYLQRTVKHPTLLQDPLDRLQFLESSELPRAVNTQALSGAGILRMVKNKAADAVNKMKTIMN 300
QY 301 EDAWFEKQOQFENLQDQRLKHLVSVYALVCHRKELSSANTAAFAKSAAMLGNSDHTAL 360
DB 301 EDAWFEKQOQFENLQDQRLKHLVSVYALVCHRKELSSANTAAFAKSAAMLGNSDHTAL 360
QY 361 SRALSOLAEVEEKIDQLHQOAFADFYMFSELLSDYRLTLAAVKGVDHRMKCKQKVEDA 420
DB 361 SRALSOLAEVEEKIDQLHQOAFADFYMFSELLSDYRLTLAAVKGVDHRMKCKQKVEDA 420
QY 421 QITLLKKREAEAKMVANKPKDQKQAKNEIREWEAKVQOQERPEQISKTIRKEVGRFEK 480
DB 421 QITLLKKREAEAKMVANKPKDQKQAKNEIREWEAKVQOQERPEQISKTIRKEVGRFEK 480

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirni L., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

```

QY 481 ERVKDFKTVIILKSLVQTOOOLIKYWEAFLEAKAIA 519
|||||
Db 481 ERVKDFKAVIILKSLVQTOOOLIKYWEAFLEAKAIA 519

RESULT 3

SNX1_MOUSE
ID SNX1_MOUSE STANDARD; PRT: 522 AA.
AC Q99N27;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 1.
GN SNX1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID-10116;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-11110793;
RA Chin L.S., Raynor M.C., Wei X., Chen H.O., Li L.;
RT "Hrs interacts with sorting Nexin 1 and regulates degradation of
epidermal growth factor receptor.";
RL J. Biol. Chem. 276:7069-7078(2001).
CC -1- FUNCTION: MAY BE INVOLVED IN SEVERAL STAGES OF INTRACELLULAR
TRAFFICKING. PLAYS A ROLE IN TARGETING LIGAND-ACTIVATED EGFR TO
THE LYSSOMES FOR DEGRADATION AFTER ENDOCYTOSIS FROM THE CELL
SURFACE AND RELEASE FROM THE GOLGI.
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF218916; AAG59616.1;
DR InterPro; IPR001683; PX.
DR InterPro; IPR005329; Sorting_nexin.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF03700; Sorting_nexin; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PS50195; PX; 1.
DR Transport; Protein transport; Golgi stack.
FT DOMAIN 143 272 PX.
SQ SEQUENCE 522 AA; 59044 MW; F0F05664087B4D24 CRC64;

Query Match 59.78; Score 1564.5; DB 1; Length 522;
Best Local Similarity 59.98; Pred. No. 3.5e-76;
Matches 318; Conservative 74; Mismatches 108; Indels 31; Gaps 7;

QY 2 AAERPPPL-----GDGKPTDFEDLDGDLFTSTVSTLTSSPSSPEPASLPAE 50
Db 10 ASERLPFPFGNDPSEGAAGSEAGSDTEGDIITGA---SRQSPKKT----- 63

QY 51 DISANSNGKPEVYVLDRE--DLFATEAEVSDSPEREILSEPSAPVTPYPTTL 108
Db 64 SLFPIKNGSKENGIEHQDQEPQDLFADATVELSLDSTQNNQ--KTMFGKTLIP----- 115

QY 109 IAPRIESKMSAPVFDREBIEEANGDIFDIIGVSDPKVGDGMNAYNAYRVTKT 168
Db 116 -HPTQETNSPKP---QPSYEELEEEDQDLDLVGVTDPKEIGDGMNAYVAYVVTQT 171

QY 169 SLSMFSKSEFSVKRFSDFLGLHLSKYLHVGVIVPAPKSIIVGMTKVKVGKEDSSS 228
Db 172 SLPMFRSQFAVRKFRSDFLGLYEKLSKHSQNGFVPPPPKSLIGMTKVKVGKEDSSS 231

QY 229 TEFEKRRRAALRYLQRTYKHKHTLLQDPLRQFLESSELPAVNTQALSAGILRMVKA 288

Db 232 AEFLEKRRRAALRYLQRTYKHKHTLLQDPLRQFLESSELPAVNTQALSAGILRMVKA 291
QY 289 ADVANKMTIKMNSDAWFEKQOQFENLDQOLKRLHVSVEALYCHRKLSANTAFKSA 348
Db 292 TDVSKMTIKMNSDIWFEKQOQFENLDQOLKRLHVSVEALYCHRKLSANTAFKSA 351
QY 349 AMLGNSDHTALSALSAEVEEKIDQLHQFQAFADFYMFSELLSDYIRLIAAVKGVFD 408
Db 352 AMLGNSDHTALSALSAEVEEKIDQLHQFQAFADFYMFSELLSDYIRLIAAVKGVFD 411
QY 409 HRMKQKQWEDAQITLLKREAEAKMVKPKIOQAKNEIREWEAKVQOQSERDFEQIS 468
Db 412 QRKMTQWQDAQITLLKREAEAKMVKPKIOQAKNEIREWEAKVQOQSERDFEQIS 471
QY 469 KTRIVGVGFERVKDKPTVILKILVLESVQTOOOLIKYWEAFLEAKAIA 519
Db 472 TVVRKEVTRFEKSKDKFNHVKYIKYLETLLHSQOOLAKYWEAFLEAKAIA 522

RESULT 4

SNX1_MOUSE
ID SNX1_MOUSE STANDARD; PRT: 522 AA.
AC Q9WV80; Q9EQZ9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 1.
GN SNX1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID-10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RA Takahara K., Omatsu Y., Maeda Y., Shimoyama S., Inaba K.;
RT "Complete sequence of mouse sorting nexin 1 (Snx1). cDNA."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Nakamura N., Wada Y., Futai M.;
RT "Mouse sorting nexin 1."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN SEVERAL STAGES OF INTRACELLULAR
TRAFFICKING. PLAYS A ROLE IN TARGETING LIGAND-ACTIVATED EGFR TO
THE LYSSOMES FOR DEGRADATION AFTER ENDOCYTOSIS FROM THE CELL
SURFACE AND RELEASE FROM THE GOLGI.
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF154120; AAD38805.1;
DR EMBL; AB019214; BAB20283.1;
DR MGD; MGI:1928395; Snx1.
DR InterPro; IPR001683; PX.
DR InterPro; IPR005329; Sorting_nexin;
DR Pfam; PF00787; PX; 2.
DR Pfam; PF03700; Sorting_nexin; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PS50195; PX; 1.
KW Transport; Protein transport; Golgi stack.
FT DOMAIN 143 272 PX.
FT CONFLICT 31 31 A -> G (IN REF. 2).
FT CONFLICT 117 117 S -> P (IN REF. 2).


```

FT CONFLICT 124 124 S -> C (IN REF. 2).
FT CONFLICT 138 138 MISSING (IN REF. 2).
FT CONFLICT 389 389 S -> F (IN REF. 2).
SQ SEQUENCE 522 AA: 58952 MW: 1EFC06B3EA551311 CRC64;

Query Match
Best Local Similarity 59.4%; Score 1557.5; DB 1; Length 522;
Matches 317; Conservative 75; Mismatches 108; Indels 31; Gaps 7;

QY 2 AAREPPPL-----GDGKPTDFELEDGEDFTSTVSTLESSPSPASPALPAE 50
Db 10 ASERLPPPPGMDPESEGAAGASEPAGSDTEGEDIFTGAAAA--TKPOSPKTT--- 63

QY 51 DISANGSKPTEVVLDDRE--DLFAEATEEVSLSDPEREPILESSEPSPAVTPVPTTL 108
Db 64 SLFPIKNGSKENGIEDQDQEPDLFADATVLSLSTQNNQ--KTMPGKTLTSHSPQ-- 119

QY 109 IAPRIESKMSAPVIFDRSREIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRYVTKT 168
Db 120 -----EATNSPKP---QPSVEELEEEQEDQDFTLVGITDPEKIGDMNAYVAYKVTQT 171

QY 169 SLGMSKSEFSVRRFSDFLGLHSLKASKYLHVGYIVPPAPEKSIYGMTKVKVGKEDSSS 228
Db 172 SLPMFRSQFAVKRRFSDFLGLYELKSEKHSQNGFIVPPPPPEKSLIGMTKVKVGKEDSSS 231

QY 229 TEFVKKRAALERYLQRTVYKHTPLQDPDLQFLESSELPRAYNTQALSAGILRWYKA 288
Db 232 AEFLEKRAALERYLQRTVYKHTPLQDPDLQFLESSELPRAYNTQALSAGILRWYKA 291

QY 289 ADVANKTKIKNESDIAWFEKQOOFENLDQRLKHLVSVYVPPAPEKSIYGMTKVKVGKEDSSS 348
Db 292 TDVANKTKIKNESDIWFEKQOOFENLDQRLKHLVSVYVPPAPEKSIYGMTKVKVGKEDSSS 351

QY 349 AMLGNSDHTALSRLSQAEEVEKIDQLHQBQAFADYMFSELSLSDYIRLIAAVKGVFD 408
Db 352 AMLGSSDHTALSRLSQAEEVEKIDQLHQBQAFADYMFSELSLSDYIRLIAAVKGVFD 411

QY 409 HRMKCQKQWEDAQITLLKKEAEAKMVKVANKPKIQOAKNEIREWEAKVOGGERDFOIS 468
Db 412 ORMKTWQWQDAQITLLKKEAEAKMVKVANKPKIQOAKNEIREWEAKVOGGERDFOIS 471

QY 469 KTRIKYGRKEKVRKFKVIKYLESLVQTOOQLIKYWEAFLEPEAKAIA 519
Db 472 TVVRKVTREKESKDFKHVHKYLESLVQTOOQLIKYWEAFLEPEAKAIA 522

RESULT 5
SNX1_HUMAN STANDARD; PRT; 522 AA.
AC Q13596; O60750; O60751;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 1.
GN SNX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=96067.
RN 1.
RP SEQUENCE FROM N.A.
RX MEDLINE=96208617; PubMed=8638121.
RA Kurten R.C.; Cadena D.B.; Gill G.N.;
RT "Enhanced degradation of EGF receptors by a sorting nexin, SNX1."
RL Science 272:1008-1010(1996).
RN 2.
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=99038232; PubMed=9819414;
RA Haft C.R., de la Luz Sierra M., Barr V.A., Haft D.H., Taylor S.I.;
RT "Identification of a family of sorting nexin molecules and
RT characterization of their association with receptors."
RL Mol. Cell. Biol. 18:7278-7287(1998).
RN 3.

```

```

409 ADQNKTKWQDQAQVLOKREAEALLWANKPKLOQAQDELEWESRVTYQTERDFE 468
466 QISTKIREVGFERRKVDFTVIKYLESLVQVQOOLIKYKWEAFLEPAKAIA 519
469 RISTVVRVIRFEKSKDFKPNHVIKYLETLLYSQOQLAKYWEAFLEPAKAIS 522

RESULT 6
VPSS_YEAST
ID VPSS_YEAST STANDARD; PRT; 675 AA.
AC 092331; 008483;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE vacuolar protein sorting-associated protein VPSS5.
GN VPSS5 OR GRD2 OR YOR069W OR YOR29-20.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97318765; PubMed=9175702;
RA Nothwehr S.F., Hinds A.E.;
RT "The yeast VPSS5/GRD2 gene encodes a sorting nexin-1-like protein
RL J. Cell Sci. 110:1063-1072(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97431786; PubMed=9285823;
RA Horzodavsky B.F., Davies B.A., Seaman M.N.J., McLaughlin S.A.,
RA Yoon S., Emr S.D.;
RT "A sorting nexin-1 homologue, Vps5p, forms a complex with Vps17p and
RT is required for recycling the vacuolar protein-sorting receptor."
RL Mol. Biol. Cell 8:1529-1541(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97279235; PubMed=9133743;
RA Valens M., Bohn C., Daignan-Fornier B., Dang V., Bolotin-Fukuhara M.;
RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals
RT the presence of two tRNAs and 24 new open reading frames."
CC -1- FUNCTION: REQUIRED FOR RETENTION OF LATE GOLGI MEMBRANE PROTEINS
CC AND VACUOLAR BIOGENESIS.
CC -1- SUBUNIT: INTERACTS WITH VPS17.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PTM: PHOSPHORYLATED ON SERINE RESIDUE(S).
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 13.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U73512; AAB62976.1;
CC EMBL; U84735; AAB41798.1;
CC EMBL; Z70678; CA94554.1; ALT_FRAME.
CC EMBL; Z74977; CA99262.1;
CC SGD; S0005595; VPSS5.
CC InterPro; IPR001683; PX.
CC Pfam; PF00787; PX; 1.
CC SMART; SM00312; PX; 1.
CC PROSITE; PS50195; PX; 1.
CC Transport; Protein transport; Golgi stack; Phosphorylation.
CC DOMAIN 279 394 PX.
CC SEQUENCE 675 AA; 76484 MW; FBA9994EADBC2BD CRC64;

```

```

Query Match 14.7%; Score 386; DB 1; Length 675;
Best Local Similarity 23.3%; Pred. No. 1.4e-13;
Matches 141; Conservative 114; Mismatches 220; Indels 130; Gaps 21;

QY 6 EPPPLGDKKPTDFEOL---EDGEDLFTSTVS-----TLESSPSSPEPA----- 45
DB 97 EODPIADLKNSTTOPATRESGALFTGNANSPLVFDDTIYDANTSPNTSISGRSGK 156
QY 46 -----SLPAEDISANSN-----GPKPTEWV---LDD----- 68
DB 157 PRILEDSARAQRNSRNLAKRTTASDDTIKTPTDPLKKAENEFVEEPLDNR 216
QY 69 --DREDLFAEATEEVSILDSPEPILSESPSPAVPTPTTLIAP-RIESKMSAPVIFD 125
DB 217 RENNEGKFTASVEKNILEQVDR-PLYN-----LPQIGANISSPAEVEENSEK---FG 264
QY 126 RSR--EETEEEAANGDIFDIETIGVSDPEKVGGMNAYMYRTTKTSLSMFSKSEFSVRR 183
DB 265 KTKIGSKVPPTTEKAVAFKVE--VKDPVKVGELTSHVEYTVISESSLLELKYAQ--VSR 320
QY 184 FSDFLGLHSLKSLKYLHVGYIVPPAPEKSIYGMVTKVKVGEDSSSTEFVEKRAALERYL 243
DB 321 YRDFRWLYRQLQNN--HWGKVIPIPPPEKQSVG-----SKENFIENRRFQMESML 368
QY 244 ORTVKHPTLLDQDLRQFLESSELPRAYNTQA-LSGAGILR----- 283
DB 369 KRICODPVLOKDKDPLFLTSDPFSESKKRAFLUGSAINDSNLSSEVRSEIQLLGA 428
QY 284 -----MVNKAADAVNK---MTI-----KMNESDAWFEKQOQFENLQQLRKLHVSVEA 329
DB 429 DAAEVLLKNGGIDAESHKGFMSISFSSLPKYNEADEFFIEKKQKIDEDNLKLSKSL 488
QY 330 LVCHRKELSAATAFA---KSAAMLGNSDHTALRSALAEVEEKIDQLHQEQAFADF 386
DB 489 VDTSENTLAASTEESFSSNVETIASLVNSEPNSL---LNNFADVHKSIKSSLSRSLQ 545
QY 387 YMFSELLSDYIRLIAAVKGVDFHRMKCQKWEDAQITLLKKREAEAKMMVANKPKIOQA 446
DB 546 LTMGVMLDDYIRSLASVKAIFNQSKLGYFLVVIENDNKKHSQLGKLGQNHSEKFFRM 605
QY 447 KNEIREWBAKVOQGERDPEIQISKIRKEVGRFERKVRKDFKTVIILKYLESLVQVQOOLIK 506
DB 606 RKEFOTLERRYNLTKQWQAVGDKIKDEFQGFQSTDKIREFRNGMEISLEAAIESQKECIE 665
QY 507 YWEAF 511
DB 666 LWETF 670

RESULT 7
SNX5_HUMAN
ID SNX5_HUMAN STANDARD; PRT; 404 AA.
AC Q9Y5X3; Q9BWP0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 5.
GN SNX5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21378165; PubMed=11485546;
RA Teasdale R.D., Locci D., Houghton F., Karlsson L., Gleeson P.A.;
RT "A large family of endosome-localized proteins related to sorting
RT nexin 1."
RL Biochem. J. 358:7-16(2001).
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;

```

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L., Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M., Chapman J.C., Clamp M., Clark A.P., Clark L.N., Clark S.Y., Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Grahman D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Lehar V.L., Martin S.L., McConachie L.J., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConachie L.J., Mclay K., McMurray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovarian carcinomas;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y., Saito K., Yamamoto Y., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Kanehori K.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May be involved in several stages of intracellular trafficking.
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
DR EMBL; AF121855; AD27828.1; -;
DR EMBL; AL121585; CAC00471.1; -;
DR EMBL; AK001793; BA91914.1; -;
DR EMBL; BC000100; -; NOT ANNOTATED_CDS.
DR Genew; HGNC:14969; SNX5.
DR MIM; 605937; -;
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PS01019; PX; 1.
KW Transport; Protein transport.
FT DOMAIN 25 172 V -> L (IN REF. 3).
FT CONFLICT 279 279
SQ SEQUENCE 404 AA; 46816 MW; 87A85620AF827EC6 CRC64;

Query Match 11.88; Score 308.5; DB 1; Length 404;
Best Local Similarity 26.48; Pred. No. 9.2e-10;
Matches 111; Conservative 79; Mismatches 161; Indels 69; Gaps 12;

QY 125 DRSR-EEIEEEANGD---IFDIEIGVSDEPKVGDGNMAYRVTKTSLSMFSKSEFSV 180
DB 14 DRSKRSVSDVLDNVDPSLQIDIPDALSERDKV-----KFTVHTKTLPTFQSPSEFSV 65
QY 181 KRPFDFGLHSLKSLKSLVHGYIIVPPAPEKSIIVG-----MTKVKVGKEDSSSTBFVEKR 235
DB 66 TRQHDFFVWLHDTLTETDYAGLIIPAPTKDFDQPREKMQKLGESEGMTKEFAKMK 125
QY 236 R-----AALERYLQRTVTKHTLQDPLQFLESSE--LPRAVNTQALSGA 279
DB 126 QLEAEYLAVFKTKTSSHEVFLQRLSSHPVLSKDRNFHFVEYDQDLSVRRKNTKEMFG- 184
QY 280 GILRMVNRKAADVNMKTIMKNESDAWFEKQOQFNDLQDLKHLVSEALVCHRKELS- 338
DB 185 GFPKSVKSADEVLFTGVK--EYDDFFEQEKNFLNYNRKIDSCVKADKMTFRSHKNVAD 242
QY 339 --ANTAFAKSAAMLGNSDHTLSALSQLAEVEEKIDQLHQEQAFADFYMFSELLSDY 396
DB 243 DYIHTAACHLSLAL---EPTVKKYLLKVAELPKLKKRGEVSSDEDLKLTELLRYV 298
QY 397 IRLIAVGVGFDRMKCKWQKWDADAQITLTKKREAEAKMVAANKPKIQAKNEIREWEAK 456
DB 299 MLNIEAAKDLLRYRTKALIDYENSKAL-----DKARLKSQDKVLAEEAH 342
QY 457 VQGERDFEQISKTIRKEVGRFEKRVKDFKTVIIKYLE-----SLVTOQOLIK 506
DB 343 QOECCQKFQOLSESAKEELINFKRRKRVAAFRKLIEMSELETKHARNNVSLQSCIDLFK 402
RESULT 8
SNX5 MOUSE
ID SNX5_MOUSE STANDARD; PRT: 404 AA.
AC Q9DB08;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 5.
GN SNX5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schraml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsch G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz K., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May be involved in several stages of intracellular trafficking (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.


```

Oy 485 DFKVVIKYLESLVOTQOOLIKYWEAF 512
Db 344 DIKLAFTDMAENIHYYEQCLATWESFL 371

RESULT 10
SNX7_MOUSE STANDARD; PRT; 387 AA.
ID SNX7_MOUSE
AC Q9CY18;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Sorting nexin 7.
GN SNX7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic liver;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boirelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RL "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- FUNCTION: May be involved in several stages of intracellular
CC trafficking (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AK011015; BAB27333.1;
DR MGD; MGI:1923811; Snx7.
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PS0195; PX; 1.
KW Transport; protein transport.
FT DOMAIN 30 151 PX.
SQ SEQUENCE 387 AA; 45000 MW; 5ED78D359CD32A1F CRC64;

Query Match 11.1%; Score 291.5; DB 1; Length 387;
Best Local Similarity 23.6%; Pred. No. 6.9e-09;
Matches 106; Conservative 76; Mismatches 149; Indels 119; Gaps 17;

Oy 97 SPVPTVPTTLPIAESKMSAPVIFDRSEIEEANGDFIDIEIGVSPKVGDCM 156
Db 7 SP-MNPTSLSMI-----NQIKFEDGPKLDFITVDAPESHVTTI 46

```

```

Oy 157 NAYMAYVTTKTSLSMFSKSEFSVKRRFSDFLGLHSLKASKYLHVGYIVPPAPEKSI VGM 216
Db 47 ETFTITVTKTSGRGEFSDSEFEVRRRIQDFLWLKGL--EEAHTLIIPLEKFI V-- 102
Oy 217 TKVKGKEDSSSTFEVKKRAALERYLQRTVKHTPLLDQPLDROFL--ESELPRAVNTQ 274
Db 103 ----KGVERFNDDFIETRRKALHKLNLRIADHPTLTFFNEDKVLTAQAEEL----SSY 154
Oy 275 ALSGAGILRMVKNKAADVANKMT-----IKMNSDAWFEKQKQFENLDQOLRK-- 322
Db 155 KKGPGILLSRMGQTVRAVASSMRGKKNRPEEFEMNNFIETFSQKINLIDKISORIYKEE 214
Oy 323 -----LHV-----SVEALVCHRKLSANTAFAKSAAMLGNSDHTALSRLSQ 366
Db 215 RDPFDEKKEVGPPIHLSASEEELVDTLKMGAGICECKA-----TEKRMAGLSEAL-- 267
Oy 367 LAEVEEKIDQLHQPAFADPYMFSELLSDYIRLIAVKGVPDHRMKCKWKWEDAQITLLK 426
Db 268 -----LPVVH-----EYVLYSEML-----VGVNKRDRDQI--QTELDSKVEALT 303
Oy 427 KREAEAKMM---VANKPDKIOQAKNEIR-EWEAKVQOGERDFEIOISKTIRKEVGREFKER 482
Db 304 YKADIDLLTTEIGKLEKDEKVCANNALKADWE-----RWKONM 341
Oy 483 VKDFKTVIILKYLESLVOTQOOLIKYWEAF 512
Db 342 KNDLSAFTDAEQONIRYVEOCLATWESFL 371

RESULT 11
SNX6_HUMAN STANDARD; PRT; 405 AA.
ID SNX6_HUMAN
AC Q9UNH7; Q9Y449;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 6 (TRAF4-associated factor 2).
GN SNX6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21378165; PubMed=11485546;
RA Teasdale R.D., Loci D., Houghton F., Karlsson L., Gleeson P.A.;
RA "A large family of endosome-localized proteins related to sorting
RT nexin 1.";
RL Biochem. J. 358:7-16(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Toji S., Yano M., Kobayashi A., Tamai K.;
RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=21276432; PubMed=11279102;
RA Parks W.T., Frank D.B., Huff C., Haft C.R., Martin J., Meng X.,
RA de Caestecker M.P., McNally J.G., Reddi A., Taylor S.I., Roberts A.B.,
RA Wang T., Lechleider R.J.;
RA "Sorting nexin 6, a novel SNX, interacts with the transforming growth
RT factor-beta family of receptor serine-threonine kinases.";
RL J. Biol. Chem. 276:19332-19339(2001).
CC -1- FUNCTION: May be involved in several stages of intracellular
CC trafficking.
CC -1- SUBUNIT: Interacts with TGFBR receptors.
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



```
QY 301 -ESDAFEERQOFENLDQOLRLHVSVEALVCHRKELSANTAFAKSAAMLG----- 352
Db 212 ADIOQAFAISRELIRNIYNSFHKLRDRAERIASRAINDAADLLIFGKELSAIGSDTTPLP 271
QY 353 -----NSEDHTALSRLASQAE-----VEEKIDQLHQEQ--AFADFVMSSELLSDYIRLI 400
Db 272 SWAALNSSTWGLKQALGLSVFALLADRAAQOQGEENDVVEKLNFLDLDLQSYKDLIC 331
QY 401 AA-VKGVDFHRMKCMOKWEDAQITLLKKREAEMKMMYANKPKDIQQAQAKNIREWEAKVQ 459
Db 332 ERHEKGVLRHQRALHKYS-----LMKQMSATAQNRPESEVEQLESRIVEQENAIQT 385
QY 460 GERDFEQISKTIRKEVGRF-----EKERVKDF-----KTVIUKYLESVOTQOOLIKYHEAF 511
Db 386 ME-----LRNYSLYCHLOETQOLIHYLPLTSHILRAFNYSQIQGHKMSVWNDL 436
QY 512 LPEAKAI 518
Db 437 RPKLSCL 443

RESULT 15
SNX9_HUMAN
AC O9Y5X1; Q9UJH6; Q9UP20; Q9BSI7; PRT; 595 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 9 (SH3 and PX domain-containing protein 1) (SDP1 protein).
GN SNX9 OR SH3PX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=21378165; PubMed=11485546;
RA Teasdale R.D.; Lodi D.; Houghton P.; Karlsson L.; Gleeson P.A.;
RT "A large family of endosome-localized proteins related to sorting nexin 1."
RL Biochem. J. 358:7-16(2001).
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=20002705; PubMed=10531379;
RA Howard L.; Nelson K.K.; Maciewicz R.A.; Blobel C.P.;
RT "Interaction of the metalloprotease disintegrins MDC9 and MDC15 with two SH3 domain-containing proteins, endophilin I and SH3PX1."
RL J. Biol. Chem. 274:31693-31699(1999).
RN [3]
RX SEQUENCE FROM N.A.
RA Zhang J.S.; Smith D.I.;
RT "Identification of differentially expressed genes in matched prostate cancer and normal epithelial cell lines."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RX SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RX SEQUENCE OF 101-549 FROM N.A.
RA Almeida J.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RX SEQUENCE OF 201-595 FROM N.A.
RA Ramanathan G.; Subramaniam V.N.; Hong W.;
RT "Human SDP1."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May be involved in several stages of intracellular trafficking.
CC -!- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
```

```
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensed@isb-sib.ch).
CC -----
DR EMBL; AF121859; AAD27832.1; -
DR EMBL; AF131214; AAF04473.1; -
DR EMBL; AF172847; AAL54871.1; -
DR EMBL; BC005022; AAH05022.1; -
DR EMBL; AL035634; CAB46196.1; -
DR EMBL; AF076957; AAD43001.1; -
DR Gene; HGNC:14973; SNX9.
DR MIM; 605952; -
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF000018; SH3; 1.
DR Pfam; PF00787; PX; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS0195; PX; 1.
DR PROSITE; PS0002; SH3; 1.
KW Transport; protein transport; SH3 domain.
FT DOMAIN 1 62
FT DOMAIN 250 361
FT CONFLICT 89
FT CONFLICT 89
FT CONFLICT 89
SQ SEQUENCE 595 AA; 66591 MW; 963892AC1A5A9227 CRC64;

Query Match 7.6%; Score 200; DB 1; Length 595;
Best Local Similarity 20.1%; Pred. No. 0.00085;
Matches 103; Conservative 81; Mismatches 213; Indels 116; Gaps 17;

QY 29 TSVSTVLESFSPSE-----PASLPAEDISANSNGPKPTVVLDDRDLEFAEATEVSLD 84
Db 138 TNPNNWDTAFGHQPAQYQGPATGDDDDWDGPKSSSYFKDSADAGAGQNGSRAS 197
QY 85 SPEREPILSSEPSAVTPTPTTLIAPRIEKSNSAPVIFDRSREIEEANGD----- 138
Db 198 SSSMKIPLNFKPGFA-KPGTEQYLLAKQAKPEKPIIV-----GDYCPMWY 244
QY 139 ----IFDIEIGVSDPEKVGD--GMNAYMAYRTTKTSLSMFSKSEFSVKRRSDFLGLHS 192
Db 245 YPTSTEDCV--VADPRKSGMYGLKSYIEQLTP-----TNTNRSVNHRYKHFDFLYE 295
QY 193 KLASKYLHVGYIVPPAPEKSIIVGKTKVVKYKEDSSSTEFVEKRAALERYLQRTVKHPTL 252
Db 296 RLLVKF-GSAIPISPLDKQVTRFE-----EEFIKMRMERLQAWNTRCRHPVI 344
QY 253 LQDPLRLQFL-----ESSELPRAVNTQALSGAGI-----LRMVNKAADAVNK 294
Db 345 SESEVFOQFLNFRDEKEWTKGRKRAERDELAGVMIFSTMEPEAPDDLDLVEIQKCEAVGK 404
QY 295 MTIKMNES-DAWFEKQOQFENLDQOLRLHVSVEALVCHRKELSANTAFAKSAAMLGN 353
Db 405 FTRAMDDGVKELLTVGQEHMKRCGTGPLPKYQKI-----GKALQSLATVFSSS-----G 453
QY 354 SEDHTALSRLASQAEVEEKIDQLHQEQAFADYMFSE-----LSDYIRLIAAVKGV 406
Db 454 YQGETDLNDAITAGKTYEETIASLVAEQPKKDLHLFLMECNHEYKGFGLGCPFDIIGTHKA 513
QY 407 FDHMKCWQKWEDAQITLLKKREAEMKMMYANKPKDIQQAQAKNIREWEAKVQGERDFEQ 466
Db 514 IE-----KVKESDKLVATSKITLQDKQNMVNR-----VSI 543
QY 467 ISKTIRKEVGRFEKERVKDFKTVIUKYLESVQ 499
```


DB 544 MSYALQAEHHHSNRIDYNSVIRLYLEQQVQ 576

Search completed: March 4, 2003, 14:59:40
Job time : 14.4234 secs

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:55:07 ; Search time 6.06686 Seconds
(without alignments)
848.711 Million cell updates/sec

Title: US-09-763-902B-6
Perfect score: 936
Sequence: 1 MGILLGLLLGLHGLTVDYGR.....PLKATSTVKSQMDWTMDMG 175

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	870	92.9	321	4	US-09-254-465A-2
2	102.5	11.0	319	1	US-08-597-495B-22
3	102.5	11.0	319	4	US-09-068-051A-22
4	102.5	11.0	319	4	US-09-336-536-67
5	102.5	11.0	319	4	US-09-254-465A-6
6	102.5	11.0	326	1	US-08-225-477B-6
7	102.5	11.0	326	5	PCT-US95-04353-6
8	101.5	10.8	270	4	US-09-254-465A-24
9	101.5	10.8	273	4	US-09-254-465A-26
10	95.5	10.2	365	2	US-08-979-424-3
11	95.5	10.2	365	4	US-09-272-496-2
12	92	9.8	303	4	US-09-651-200-23
13	92	9.8	309	2	US-08-456-104-4
14	92	9.8	309	3	US-08-479-744A-23
15	92	9.8	309	4	US-08-280-757B-23
16	92	9.8	309	4	US-08-205-697A-21
17	92	9.8	309	4	US-08-702-525-21
18	92	9.8	309	4	US-08-651-200-22
19	92	9.8	309	5	PCT-US95-02576-21
20	92	9.8	314	4	US-08-205-697A-13
21	92	9.8	314	4	US-08-702-525-13
22	92	9.8	314	5	PCT-US95-02576-13
23	92	9.8	589	2	US-08-724-394A-1
24	90.5	9.7	365	4	US-08-928-383B-2
25	89	9.5	205	4	US-09-462-270-4
26	89	9.5	299	4	US-09-188-930-189
27	89	9.5	299	4	US-09-188-930-331

28	89	9.5	299	4	US-09-462-270-2	Sequence 2, Appl
29	89	9.5	299	4	US-09-254-465A-1	Sequence 1, Appl
30	88	9.4	205	1	US-07-870-029-2	Sequence 2, Appl
31	88	9.4	205	1	US-08-233-005-2	Sequence 2, Appl
32	88	9.4	205	1	US-08-428-943-2	Sequence 2, Appl
33	88	9.4	205	3	US-09-016-649-2	Sequence 2, Appl
34	88	9.4	205	3	PCT-US95-04858-2	Sequence 2, Appl
35	87.5	9.3	294	6	5260223-1	Patent No. 5260223
36	87	9.3	624	2	US-08-642-406A-22	Sequence 22, Appl
37	87	9.3	624	4	US-09-199-534-22	Sequence 22, Appl
38	87	9.3	624	4	US-09-199-534-22	Sequence 22, Appl
39	87	9.3	773	3	US-08-434-000A-2	Sequence 2, Appl
40	87	9.3	773	4	US-09-312-157-2	Sequence 2, Appl
41	86.5	9.2	303	4	US-08-985-950-2	Sequence 2, Appl
42	86.5	9.2	365	4	US-08-928-383B-23	Sequence 23, Appl
43	86.5	9.2	365	4	US-08-928-383B-24	Sequence 24, Appl
44	86.5	9.2	365	4	US-08-928-383B-26	Sequence 26, Appl
45	85.5	9.1	282	4	US-09-404-879A-393	Sequence 393, App

ALIGNMENTS

RESULT 1
US-09-254-465A-2
; Sequence 2, Application US/09254465A
; Patent No. 6410708
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Fong, Sherman
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Napier, Mary A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: COMPOUNDS, COMPOSITIONS AND METHODS FOR THE TREATMENT
; FILE OF INVENTION: OF DISEASES CHARACTERIZED BY A33- RELATED ANTIGENS
; FILE REFERENCE: P1216R1(US)
; CURRENT APPLICATION NUMBER: US/09/254,465A
; CURRENT FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: PCT/US98/24855
; PRIOR FILING DATE: 1998-11-20
; PRIOR APPLICATION NUMBER: US 60/066,364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 60/078,936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: PCT/US98/19437
; PRIOR FILING DATE: 1998-09-17
; NUMBER OF SEQ ID NOS: 30
; SEQ ID NO 2
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-254-465A-2

Query Match	92.9%	Score 870;	DB 4;	Length 321;
Best Local Similarity	64.7%	Pred. No. 9.9e-84;		
Matches 174;	Conservative	0;	Mismatches 1;	Indels 94;
Gaps	1;			
Qy	1	MGILLGLLLGLHGLTVDYGRFILEVPSVTGPWKGDVNLCTYDPLQGYTOVLVKWLVOR	60	
Db	1	MGILLGLLLGLHGLTVDYGRFILEVPSVTGPWKGDVNLCTYDPLQGYTOVLVKWLVOR	60	
Qy	61	GSDPVTIFLRSSGDHIOQAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVETWTP	120	
Db	61	GSDPVTIFLRSSGDHIOQAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVETWTP	120	
Qy	121	DGNQVVRKITELRVQ-----	136	
Db	121	DGNQVVRKITELRVQ-----	180	
Qy	137	-----	146	

RESULT 7
 PCT-US95-04353-6
 ; Sequence 6, Application PC/TUS9504353
 ; GENERAL INFORMATION:
 ; APPLICANT: Susan Hockfield
 ; APPLICANT: Diane M. Jaworski
 ; TITLE OF INVENTION: BEHAB, A Brain Hyaluronan-Binding Protein
 ; TITLE OF INVENTION: Luronan-Binding Protein
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: St. Onge Steward Johnston & Reens
 ; STREET: 986 Bedford Street
 ; CITY: Stamford
 ; STATE: CT
 ; COUNTRY: United States
 ; ZIP: 06905
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" 1.44 Mb diskette
 ; COMPUTER: IBM PC
 ; OPERATING SYSTEM: MS DOS
 ; SOFTWARE: Word Processor
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/04353
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/225,477
 ; FILING DATE: April 8, 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mary M. Krinsky
 ; REGISTRATION NUMBER: 32423
 ; REFERENCE/DOCKET NUMBER: 1751-P0004
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 203-324-6155
 ; TELEFAX: 203-327-1096
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 326 residues
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: polypeptide
 ; DESCRIPTION: functional domains
 ; FRAGMENT TYPE:
 ; FEATURE:
 ; NAME/KEY: rat link protein
 ; PUBLICATION INFORMATION:
 ; AUTHORS: Dose, K., Hassell, J.R., Ca-
 ; terson, B., and Yamada, Y.
 ; TITLE: Link protein cDNA sequence reveals a
 ; tandemly repeated protein sequence.
 ; JOURNAL: Proc. Natl. Acad. Sci. USA
 ; VOLUME: 83
 ; PAGES: 3761-3765
 ; DATE: 1986
 ; PCT-US95-04353-6

Query Match 11.0%; Score 102.5; DB 5; Length 326;
 Best Local Similarity 29.1%; Pred. No. 0.0083;
 Matches 37; Conservative 21; Mismatches 52; Indels 17; Gaps 7;
 QY 12 HLVDYGRILEVPES-VTGPKGQVNLPCY--DPL---QGYTVLVKLVQSGDPV 65
 Db 5 HQAENGPRLLVEAQAKVFSHGGNVTLECKFYRDPATFGSGIHKIRIKW-TKLTS-- 61
 QY 66 TIFLRD-----SSGDHQQ-QAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVTWQT 119
 Db 62 --YLRVDVVFVSGYHKKTYGGYQGRVFLKGGSDNDASLIITLTDLEDYGRYKCEVIEGL 119
 QY 120 PDGNQV 126
 Db 120 EDDRAV 126

RESULT 8
 US-09-254-465A-24
 ; Sequence 24, Application US/09254465A
 ; Patent No. 6410708
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi J.
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Napier, Mary A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: COMPOUNDS, COMPOSITIONS AND METHODS FOR THE TREATMENT
 ; TITLE OF INVENTION: OF DISEASES CHARACTERIZED BY A33- RELATED ANTIGENS
 ; FILE REFERENCE: P1216R1(US)
 ; CURRENT APPLICATION NUMBER: US/09/254,465A
 ; CURRENT FILING DATE: 1999-03-05
 ; PRIOR APPLICATION NUMBER: PCT/US98/24855
 ; PRIOR FILING DATE: 1998-11-20
 ; PRIOR APPLICATION NUMBER: US 60/066,364
 ; PRIOR FILING DATE: 1997-11-21
 ; PRIOR APPLICATION NUMBER: US 60/078,936
 ; PRIOR FILING DATE: 1998-03-20
 ; PRIOR APPLICATION NUMBER: PCT/US98/19437
 ; PRIOR FILING DATE: 1998-09-17
 ; NUMBER OF SEQ ID NOS: 30
 ; SEQ ID NO 24
 ; LENGTH: 270
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-254-465A-24
 Query Match 10.8%; Score 101.5; DB 4; Length 270;
 Best Local Similarity 27.8%; Pred. No. 0.0081;
 Matches 35; Conservative 25; Mismatches 57; Indels 9; Gaps 6;
 QY 13 LTVDTYGRILEVPESVTGPKWG-DVNLPCYDPLQGYTVLVKWN-LVQSGDPVTIFL 69
 Db 3 VTVDALS---VETPDVLRASQGSVTLPTCTYHTSTSRGLQWDKLLTHTERVIV- 58
 QY 70 ROSSGDHIQQAQYQGRHLVSHKV-PCGVSLQSLTLEMDRSHYTCVTWQTP-DGNQVVR 127
 Db 59 PFSKNYIHGELYKRVISNNNAEQSDASITDQLIMADNGTYECSVSLMSDLEGNTKSR 118
 QY 128 DKITEL 133
 Db 119 VRLVL 124
 RESULT 9
 US-09-254-465A-26
 ; Sequence 26, Application US/09254465A
 ; Patent No. 6410708
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi J.
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Napier, Mary A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: COMPOUNDS, COMPOSITIONS AND METHODS FOR THE TREATMENT
 ; TITLE OF INVENTION: OF DISEASES CHARACTERIZED BY A33- RELATED ANTIGENS
 ; FILE REFERENCE: P1216R1(US)
 ; CURRENT APPLICATION NUMBER: US/09/254,465A
 ; CURRENT FILING DATE: 1999-03-05
 ; PRIOR APPLICATION NUMBER: PCT/US98/24855
 ; PRIOR FILING DATE: 1998-11-20
 ; PRIOR APPLICATION NUMBER: US 60/066,364
 ; PRIOR FILING DATE: 1997-11-21

PRIOR APPLICATION NUMBER: US 60/078,936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: PCT/US98/19437
PRIOR FILING DATE: 1998-09-17
NUMBER OF SEQ ID NOS: 30
SEQ ID NO 26
LENGTH: 273
TYPE: PRT
ORGANISM: Homo sapiens
US-09-254-465A-26

Query Match 10.8%; Score 101.5; DB 4; Length 273;
Best Local Similarity 27.8%; Pred. No. 0.0082;
Matches 35; Conservative 25; Mismatches 57; Indels 9; Gaps 6;
QY 13 LVVDYGRILEVPESVTGPKWG-DYNLCTYDPLOGYQVLYKW--LVQSGDPVTIFL 69
DB 6 VVVDAIS---VETPDQVLRASQCKSVTLCTYHTSSREGLIQWDLKLLTHRTVVIN- 61
QY 70 RDSSGDHIQQAQYGRHLVSHKV-PGDVSLQLSTLEMDRSHYTCVETWQTP-DGNQVVR 127
DB 62 PFSKNYIHGELYKNRVSISNNAEQSDASITIDLTMDANGYECVSLMSDLEGNTKSR 121
QY 128 DKITEL 133
DB 122 VRLVL 127

RESULT 10
US-08-979-424-3
Sequence 3, Application US/08979424
Patent No. 5942606
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VIRAL RECEPTOR PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Forter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
FILING DATE: Filed Herewith
APPLICATION NUMBER: US/08/979,424
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0405 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1946351

US-08-979-424-3

Query Match 10.2%; Score 95.5; DB 2; Length 365;
Best Local Similarity 26.0%; Pred. No. 0.054;
Matches 53; Conservative 28; Mismatches 80; Indels 43; Gaps 13;
QY 1 MGILLGLLLGLHLYTDYGRPI-LEVPESVTGPKGD-VNLPCTY-----DPLQGYTQ 51
DB 1 MALLLCFVLL--CGVVDFAFARLSITTPPEMIEKAKETAYLPCKFTLSPEDQGPLD---- 54
QY 52 VLVKWLIVQRGS-----DPVTIFLRDSSGDHIQQAQY---QGRHLVSHK--VPGDVSILQLST 102
DB 55 --IEWLISPADNQKVQVILY---SGDKIYDDYYPDLKGRVHFTSNDLKS GDASINVTN 109
QY 103 LEMDRSHYTCVETWQTPDGNQ-----VVRDKITELRVQKHSSKLL-----KTRTEAPTWT 154
DB 110 LQLSDIGTYQCKVKKAPGVANKIHLVLPKPSGARYVDGSEEGSDPKICE-PREGS 168
QY 155 YPL-----KATSTVKOSWDWTTDM 173
DB 169 LPLOYEWQKLSDSQKMPSTSLAEM 192

RESULT 11
US-09-272-496-2
Sequence 2, Application US/09272496
Patent No. 6245966
GENERAL INFORMATION:
APPLICANT: Degregori, James
TITLE OF INVENTION: Adenoviral mediated gene transfer into lymphocytes
FILE REFERENCE: 90-98
CURRENT APPLICATION NUMBER: US/09/272,496
CURRENT FILING DATE: 1999-03-19
EARLIER APPLICATION NUMBER: US 60/092782
EARLIER FILING DATE: 1998-07-14
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 365
TYPE: PRT
ORGANISM: Homo sapiens
US-09-272-496-2

Query Match 10.2%; Score 95.5; DB 4; Length 365;
Best Local Similarity 26.0%; Pred. No. 0.054;
Matches 53; Conservative 28; Mismatches 80; Indels 43; Gaps 13;
QY 1 MGILLGLLLGLHLYTDYGRPI-LEVPESVTGPKGD-VNLPCTY-----DPLQGYTQ 51
DB 1 MALLLCFVLL--CGVVDFAFARLSITTPPEMIEKAKETAYLPCKFTLSPEDQGPLD---- 54
QY 52 VLVKWLIVQRGS-----DPVTIFLRDSSGDHIQQAQY---QGRHLVSHK--VPGDVSILQLST 102
DB 55 --IEWLISPADNQKVQVILY---SGDKIYDDYYPDLKGRVHFTSNDLKS GDASINVTN 109
QY 103 LEMDRSHYTCVETWQTPDGNQ-----VVRDKITELRVQKHSSKLL-----KTRTEAPTWT 154
DB 110 LQLSDIGTYQCKVKKAPGVANKIHLVLPKPSGARYVDGSEEGSDPKICE-PREGS 168
QY 155 YPL-----KATSTVKOSWDWTTDM 173
DB 169 LPLOYEWQKLSDSQKMPSTSLAEM 192

RESULT 12
US-09-651-200-23
Sequence 23, Application US/09651200
Patent No. 6429303
GENERAL INFORMATION:
APPLICANT: Green et al
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
LYMPHOCYTE ACTIVATION ANTIGEN B-7 FAMILY AND
POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-562 (CURA-62)
CURRENT APPLICATION NUMBER: US/09/651,200
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/152383
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/172909
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/183578
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 23
LENGTH: 303
TYPE: PRT
ORGANISM: Mus musculus
US-09-651-200-23

Query Match 9.8%; Score 92; DB 4; Length 303;
Best Local Similarity 24.0%; Pred. No. 0.096;
Matches 40; Conservative 33; Mismatches 76; Indels 18; Gaps 6;
QY 5 LGLLLGLHGLVDTYGRPILEVPESVT----GPKWGDVNLPCYDPLQ--GYTVLVKWL 58
DB 1 MGLAIFLVIV-----LLISDAVSVEYQAYFNGTAYLPCPFTKAQNISLSLVVWQD 53
QY 59 QRGSDPVTIFLRDSSGDHIQAKYQGRHLVSHKVPDVSLSLSTLEMDRSHYTCVETWQ 118
DB 54 QOKLVLYEHLVGTKEKLSV-NAKYLGRTSFDR---NNWTLRLHNQKDMGSDYCFIQKK 109
QY 119 TPDGNVVRDKITELRVQKHSKLLKTKTEAPTMTYPLKATSVKQ 165
DB 110 PPTGSIILQOQLTSLSVIANFSE-PEIKLAQNVGTSGINLTCTSKQ 155

RESULT 13
US-08-456-104-4
Sequence 4, Application US/08456104
Patent No. 5861310
GENERAL INFORMATION:
APPLICANT: Freeman, Gordon J.
APPLICANT: Nadler, Lee M.
APPLICANT: Gray, Gary S.
TITLE OF INVENTION: TUMOR CELLS MODIFIED TO EXPRESS B7-2 AND B7-3 WITH INCREASED
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,104
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/101,624;
FILING DATE: 26-JUL-1993;
APPLICATION NUMBER: 08/109,393;
APPLICATION NUMBER: 19-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: RPI-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 4;

SEQUENCE CHARACTERISTICS:
LENGTH: 309 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-456-104-4

Query Match 9.8%; Score 92; DB 2; Length 309;
Best Local Similarity 24.0%; Pred. No. 0.099;
Matches 40; Conservative 33; Mismatches 76; Indels 18; Gaps 6;
QY 5 LGLLLGLHGLVDTYGRPILEVPESVT----GPKWGDVNLPCYDPLQ--GYTVLVKWL 58
DB 7 MGLAIFLVIV-----LLISDAVSVEYQAYFNGTAYLPCPFTKAQNISLSLVVWQD 59
QY 59 QRGSDPVTIFLRDSSGDHIQAKYQGRHLVSHKVPDVSLSLSTLEMDRSHYTCVETWQ 118
DB 60 QOKLVLYEHLVGTKEKLSV-NAKYLGRTSFDR---NNWTLRLHNQKDMGSDYCFIQKK 115
QY 119 TPDGNVVRDKITELRVQKHSKLLKTKTEAPTMTYPLKATSVKQ 165
DB 116 PPTGSIILQOQLTSLSVIANFSE-PEIKLAQNVGTSGINLTCTSKQ 161

RESULT 14
US-08-479-744A-23
Sequence 23, Application US/08479744A
Patent No. 6084067
GENERAL INFORMATION:
APPLICANT: Freeman, Gordon J.
APPLICANT: Nadler, Lee M.
APPLICANT: Gray, Gary S.
TITLE OF INVENTION: No. 6084067el CTLA4/CD28 Ligands and
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,744A
FILING DATE: June 7, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/280,757
FILING DATE: 26-JUL-1994
APPLICATION NUMBER: 08/109,393
FILING DATE: 28-AUG-1993
APPLICATION NUMBER: 08/101,624
FILING DATE: 26-JULY-1993
APPLICATION NUMBER: 08/147,773
FILING DATE: 3-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: RPI-004CP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 23;
SEQUENCE CHARACTERISTICS:
LENGTH: 309 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-479-744A-23

Query Match 9.88; Score 92; DB 3; Length 309;
Best Local Similarity 24.08; Pred. No. 0.099;
Matches 40; Conservative 33; Mismatches 76; Indels 18; Gaps 6;

QY 5 LGLLLGLHLTVDTYGRPILEVPESVT-----GPWKGDVNLPCYDPLQ--GYTQVLVKWLV 58
Db 7 MGLAILIFVTV-----LLISDAVSVEQAYFNGTAYLPCPFTKAQNISLSLVVFWQD 59
QY 59 QRGSDPVTIFLRDSSGDHIQOAKYQGRHLVSHKVPDVSLSLSTLEMDRDRSHYTCVETWQ 118
Db 60 QOKLVLYEHYLGTEKLDV--NAKYLGRTSFDR---NNTWTLRLHNQIKDMGSDYCFIOKK 115
QY 119 TPDGNQVVRDKITELRVOKHSHKLLKTTEAPTMTYPLKATSTVKQ 165
Db 116 PPTGSIILQOQLTELSVIANESE-PEIKLAQNVGTGNSGINTCTSKQ 161

RESULT 15

US-08-280-757B-23
; Sequence 23, Application US/08280757B
; Patent No. 6130316
; GENERAL INFORMATION:
; APPLICANT: Freeman, Gordon J.
; APPLICANT: Nadler, Lee M.
; APPLICANT: Gray, Gary S.
; APPLICANT: Greenfield, Edward
; TITLE OF INVENTION: No. 6130316el CTLA4/CD28 Ligands and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,757B
; FILING DATE: 26-JUL-1994
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/101,624
; FILING DATE: 26-JULY-1993
; APPLICATION NUMBER: 08/109,393
; FILING DATE: 19-AUG-1993
; APPLICATION NUMBER: 08/147,773
; FILING DATE: 3-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: RPI-004CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-280-757B-23

Query Match 9.88; Score 92; DB 4; Length 309;
Best Local Similarity 24.08; Pred. No. 0.099;
Matches 40; Conservative 33; Mismatches 76; Indels 18; Gaps 6;

QY 5 LGLLLGLHLTVDTYGRPILEVPESVT-----GPWKGDVNLPCYDPLQ--GYTQVLVKWLV 58
Db 7 MGLAILIFVTV-----LLISDAVSVEQAYFNGTAYLPCPFTKAQNISLSLVVFWQD 59
QY 59 QRGSDPVTIFLRDSSGDHIQOAKYQGRHLVSHKVPDVSLSLSTLEMDRDRSHYTCVETWQ 118
Db 60 QOKLVLYEHYLGTEKLDV--NAKYLGRTSFDR---NNTWTLRLHNQIKDMGSDYCFIOKK 115
QY 119 TPDGNQVVRDKITELRVOKHSHKLLKTTEAPTMTYPLKATSTVKQ 165
Db 116 PPTGSIILQOQLTELSVIANESE-PEIKLAQNVGTGNSGINTCTSKQ 161

Search completed: March 4, 2003, 15:06:24
Job time : 8.06686 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:50:41 ; Search time 41.007. Seconds
(without alignments)
1559.740 Million cell updates/sec

Title: US-09-763-902B-1
Perfect score: 2459
Sequence: 1 MRFVVALVLLNVAAGAVPL.....YIPLEKDERHOWIVLLSFQL 480

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470.

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_101002.*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
 - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2459	100.0	480	21	Human protein tran
2	1205.5	49.0	251	22	Human hrgn48 prote
3	304	12.4	1072	23	Mouse ischaemic co
4	302	12.3	1021	22	Staphylococcus aur
5	302	12.3	1021	22	Staphylococcus aur
6	285.5	11.6	1802	21	Cell wall protein
7	285.5	11.6	1802	21	Staph. epidermidis
8	284	11.5	933	21	Staphylococcus aur
9	284	11.5	933	22	Staphylococcus aur
10	283	11.5	918	20	S. aureus C1FB pro

11	281	11.4	936	18	AAW89801	Staphylococcus aur
12	272.5	11.1	617	22	ABB35445	Peptide #2951 enco
13	272.5	11.1	617	22	ABB20884	Protein #2883 enco
14	272.5	11.1	617	22	AAW56269	Human brain expres
15	272.5	11.1	617	22	AAW16458	Peptide #2892 enco
16	272.5	11.1	617	22	AAW04187	Peptide #2869 enco
17	265.5	10.8	1026	22	AAW78825	Human protein SEQ
18	261.5	10.6	1033	22	AAW78809	Human protein SEQ
19	259	10.5	1664	19	AAW43106	C. thermocellum O1
20	250	10.2	1253	21	ABP19772	Human dentin sialo
21	250	10.2	1253	23	ABP51785	Human dentin sialo
22	248	10.1	654	22	ABB63266	Drosophila melanog
23	246	10.0	688	22	ABB63269	Drosophila melanog
24	242.5	9.9	1092	19	AAW41602	Staphylococcus epi
25	238.5	9.7	2112	22	ABB60403	Drosophila melanog
26	236.5	9.6	554	19	AAW20763	Human neurofilamen
27	232	9.4	699	22	AAW79156	Human protein SEQ
28	230	9.4	718	22	AAW80140	Drosophila melanog
29	227.5	9.3	931	22	ABB61093	Drosophila melanog
30	227.5	9.3	2703	22	ABB60074	Drosophila melanog
31	221	9.0	1151	22	ABB61598	Drosophila melanog
32	219.5	8.9	1071	22	ABB66158	Drosophila melanog
33	219.5	8.9	1071	22	ABB67038	Drosophila melanog
34	219.5	8.9	2858	22	ABB71150	Drosophila melanog
35	219.5	8.9	3060	22	ABB58064	Drosophila melanog
36	219	8.9	754	19	AAW33811	Tat stimulatory fa
37	217	8.8	1192	22	ABG02038	Novel human diagno
38	216	8.8	2344	22	AAU37120	Staphylococcus aur
39	215	8.7	768	22	AAW78886	Human protein SEQ
40	215	8.7	1163	22	AAU28028	Novel human secret
41	213.5	8.7	2768	22	ABB68397	Drosophila melanog
42	213	8.7	767	22	AAW79670	Human protein SEQ
43	212	8.6	1349	22	AAU34402	Staphylococcus aur
44	212	8.6	1349	22	AAU37344	Staphylococcus aur
45	211.5	8.6	488	21	AAW42093	Human ORFX ORF1857

ALIGNMENTS

RESULT 1
AAW82317

ID AAY82317_standard; Protein; 480 AA.

AC AAY82317;

DT 19-JUN-2000 (first entry)

DE Human protein transport molecule (PTAM). SEQ ID NO:1.

Human; protein transport molecule; PTAM; diagnosis; cytostatic; antiarthritic; antidiabetic; immunosuppressant; antiarteriosclerotic; antiallergic; antidiabetic; antilipase; antirheumatic; osteopathic; dermatological; antianemic; antipsoriatic; hepatotropic; antitumor; antiinflammatory; antiHIV; protein transport regulator; cancer; immune disorder; cell proliferative disorder; secretory disorder; urticaria; allergy; abnormal vesicle trafficking; asthma; autoimmune haemolytic anaemia.

OS Homo sapiens.

PN WO200012703-A2.

XX 09-MAR-2000.

XX 26-AUG-1999; 99WO-US19616.

XX 27-AUG-1998; 98US-0098206.

XX (INCY-) INCYTE PHARM, INC.

XX Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;

PI Gorgone GA, Baughn MR, Patterson C;

RESULT 2

SQ Sequence 1021 AA;
 Query Match 12.3%; Score 302; DB 22; Length 1021;
 Best Local Similarity 22.3%; Pred No. 8.e-12;
 Matches 84; Conservative 81; Mismatches 188; Indels 24; Gaps
 QY 22 ATSVKQEEAGVRSAGNVSTHPSLSORPGGS-----TKSHPEPQTPKDSPSKSSAEA 74
 Db 606 ASDSDTSDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDS 665
 QY 75 QTPEPTPNKGAEAKTQKDNSSNKGSGAEATKGTGSKSGSEAAQTTKDVPNKSGADGOTPKDGSS 74
 Db 666 DSD 725
 QY 135 DSTGKGAEAQTPEDSPNRSGAEAKTKDPSKSGSEAAQTTKDVPNKSGADGOTPKDGSS 194
 Db 726 DSD 785
 QY 195 KSGAEOQTPKDVPNKGAEKQTPKDGNKSGAEEOGPIDGPKSGAEQTSKDSPNKYVP 254
 Db 786 DSD 845
 QY 255 EQPSRKDHKSIPNSDNKELPKADTNQLADKGLSPHAFKTESGEETDLISPPOEYKS 314
 Db 846 ESD 896
 QY 315 SEPTDEVPEKEADDGTGPEEGSPPKEEKKMGSGAS---SENREGTLSDS-----TGSE 366
 Db 897 SDSDSGSDSASDSGSDSGSN 956
 QY 367 KDDLYPNGSGNGSAESS 383
 Db 957 NNVPNPSPKNGTNASN 973
 RESULT 5
 AAU36951
 ID AAU36951 standard; Protein; 1021 AA.
 AC AAU36951;
 DT 14-FEB-2002 (first entry)
 XX Staphylococcus aureus cellular proliferation protein #1121.
 DE Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 OS Staphylococcus aureus.
 XX WO200170955-A2.
 XX 27-SEP-2001.
 PF 21-MAR-2001; 2001WO-US09180.
 XX 21-MAR-2000; 2000US-191078P.
 PP 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207272P.
 PR 23-OCT-2000; 2000US-243578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KL, Zyskind JW, wall D, Trawick JD, Carr GJ;
 PFI Yamamoto RT, Xu HH;
 PI WPI: 2001-611495/70.
 DR N-FSDB; AAS54810.
 DR New polynucleotides for the identification and development of
 XX

XX The patent discloses multicomponent vaccines containing selected
 CC combinations of bacterial binding proteins termed MSCRAMM (microbial
 CC surface components recognising adhesive matrix molecules) or their
 CC antibodies. A vaccine composition is provided that includes collagen
 CC binding protein or peptide, e.g. CNA, a fibrinogen binding protein
 CC preferably Clumping factor A (ClfA), a clumping factor B (ClfB),
 CC and optionally a fibronectin binding protein e.g. FnBP-A.
 CC The vaccines are useful for imparting protection against a broad
 CC spectrum of Staphylococcal strains and for inhibiting microbial
 CC colonisation, especially of Staphylococcus aureus, in an animal.
 CC The combinations can also be used to select donor blood pools for the
 CC preparation of purified blood products for passive immunisation.
 CC The present sequence is a serine-aspartate repeat region
 CC protein, SdrF from Staphylococcus epidermidis. The Sdr protein is
 CC useful in vaccine preparation in combination with specific
 CC bacterial binding proteins. These vaccines can be used to treat a broad
 CC spectrum of bacterial infections, including those arising from both
 CC coagulase-positive and coagulase-negative bacteria.
 XX Sequence 1802 AA;

Query Match 11.6%; Score 285.5; DB 21; Length 1802;
 Best Local Similarity 22.0%; Pred. No. 2e-10;
 Matches 93; Conservative 84; Mismatches 209; Indels 37; Gaps 8;

QY 56 SHPEPQPKSPKSSAEAOPTDPNKGAEAKTKQDSSNKGAEAKTKGTSKSGSE 115
 Db 1411 SDSDSDDADSD 1470
 QY 116 AQTGKSTKSHSELOTPKDTGKSGAEAOPTDPNKGAEAKTKQDSPKSGSEAO 175
 Db 1471 SDADSD 1530
 QY 176 KDVPNKGAGDQTPKGGSKGAEADQTPKVPNKGAEAKTKQDSSNKGAEAGQPIDGP 235
 Db 1531 SDSDSDDSD 1590
 QY 236 SKSGAEQTSKSDPNKVPQPKSKHSPKPSNPNKELPKADTNLQKGLSPHAFK 295
 Db 1591 SDSDSDDSD 1650
 QY 296 TEGSEETDLISPPQEEVKS-----SEPTDEVPEKAEADDDTGPEEGSPKPEEKMSGSA 350
 Db 1651 SDSDSDDADSD 1710
 QY 351 S-SENREGTLDSTGSEKDDLYPNGSNGSGAESHHFFAYLVTAAILVAVLYIAHNNRK 409
 Db 1711 SDSDSDDSD 1753
 QY 410 IAFVLEGRKSKVTRPRKASDYQRLDQKYLILNVFPAPPKRSLPQVLTWYIPLEKDER 469
 Db 1754 LGALLGRRRRKKONKEX-----HYXFIHLFQA---RSIWPLGLKS---XYIERRRK 1799
 QY 470 HQW 472
 Db 1800 XWV 1802

RESULT 8
 AAY58435
 ID AAY58435 standard; Protein; 933 AA.

XX AAY58435;
 XX
 XX 27-MAR-2000 (first entry)
 DT Staphylococcus aureus fibrinogen binding ClfA protein.
 DE
 XX ClfA; fibrinogen binding protein; bacterial colonisation;
 XX indwelling medical device; staphylococcal infection.
 XX
 OS Staphylococcus aureus.

XX Key Location/Qualifiers
 FH Peptide 1..39
 FT /note= "Signal peptide"
 FT Region 40..559
 FT /note= "Region A"
 FT Region 332..550
 FT /note= "Fibrinogen-binding region"
 FT Region 560..867
 FT /note= "Region R"
 FT Region 896..900
 FT /note= "Gram positive wall-associated consensus motif"
 XX US6008341-A.
 PN 28-DEC-1999.
 XX 22-AUG-1994; 94US-0293728.
 XX 22-AUG-1994; 94US-0293728.
 XX (QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.
 PA Foster TJ, McDevitt DL;
 PI WPI; 2000-096389/08.
 DR N-PSDB; AAY55832.
 XX Nucleic acid encoding the fibrinogen-binding protein of Staphylococcus aureus, useful for treatment or prevention of infections
 PT Disclosure; Fig 2A-1-4; 35pp; English.
 XX This sequence represents the Staphylococcus aureus
 CC fibrinogen-binding protein, ClfA. ClfA is an important receptor
 CC involved in S. aureus colonisation of indwelling medical devices (e.g.,
 CC catheters, artificial heart valves). Shortly after implantation, the
 CC surfaces of medical devices become coated with host plasma and matrix
 CC proteins such as fibrinogen and fibronectin, and there is considerable
 CC evidence to suggest that bacterial adherence to fibrinogen/fibrin is
 CC important in the initiation of device-related infection. The
 CC fibrinogen-binding region of ClfA is thought to reside between residues
 CC 332 and 550 in a region designated A. The protein also contains a
 CC repeated region (region R) comprising 154 repeats of the dipeptide
 CC Ser-Asp, and the C-terminus contains features present in surface
 CC proteins of other Gram positive bacteria that are responsible for
 CC anchoring the protein to the cell wall and cell membrane. ClfA, or its
 CC fragments, may be used to block S. aureus colonisation of wounds, to
 CC prevent adherence of S. aureus to indwelling medical devices, as
 CC vaccines to protect against S. aureus infection (e.g., mastitis in
 CC ruminants), to raise specific antibodies, and for diagnosis (by
 CC agglutination or immunoassay). The specific antibodies
 CC are used for passive immunisation, to block infection of wounds or
 CC adhesion of S. aureus and for diagnosis. Nucleotides encoding ClfA and
 CC its fragments may be used as diagnostic probes.
 XX Sequence 933 AA;

Query Match 11.5%; Score 284; DB 21; Length 933;
 Best Local Similarity 22.1%; Pred. No. 1.2e-10;
 Matches 80; Conservative 83; Mismatches 173; Indels 26; Gaps 5;

QY 37 AGNVSTHPSLSQRPQ--GSTKSHPE-----PQTPKSPKSSAEAOPTDPNKGAEAK 89
 Db 535 SGBGIDKPVVYPEQDEPGEIPEIPEDSDSDPGSDGSDNSDSDSGSDSTSDSGSDSA 594
 QY 90 TOKDSSNKGAEAKTKGTSKSGSEAOPTKDTKSHSELOTPKDTGKSGAEAOPTDP 149
 Db 595 SDSDSASDSDASDSDSDASDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 654
 QY 150 SPNRSGAEAKTKQDSPKSGSEAOPTKDPNKGAGDQTPKDGSSKSGAEADQTPKDP 209
 Db 655 SDSDSDDSD 714

535	SCGDIDKPVVPEQDPCEIEPIPEDSDSDPCSDSGSDSNDSGSDSGSDSTSDSGSDSA	594
90	TQKDSNKGABAKTQKSTSGSEAOQTCKSHSELQTPKDSGTGKSGAEAOPTED	149
595	SDSDASDSDASDSDASDSDASDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD	654
150	SPNRSAGAEAKTKDSPKSGSEAOQTKVPKSGAGDQTPKDGSKSGAEDQTPKDPNK	209
655	SD	714
210	SGAEKQTPKDGSKSGAEQIQIDGPKSGAEQTSKDSFNKVPPEQFSRKDHKSPISNP	269
715	SD	774
270	SNKELPKADTQNLADKGLSPHAKTESGETDLISPPQEEVKS-----SEPTEDVE	322
775	SPSD	834
323	PREAEDDDTGPPEGSPPEKEKMKSGSASSENREGTFLSDS-TCSEKDDILYPNGSGNGSAE	381
835	SDSASDSDSGSDSDS-----SSDSDSESDSDSDSESGSNNNVPPNSPKNGTNA	893
382	SS 383	
884	SN 885	
RESULT 10		
AY08640		
ID	AY08640 standard; protein; 918 AA.	
XX	AY08640;	
AC		
XX		
XX	09-AUG-1999 (first entry)	
XX		
DE	S: aureus ClfB protein.	
XX		
KW	Fibrinogen-binding protein; alpha chain; beta chain; ClfB; SdrC; SdrD;	
KW	SdrE; fibrinogen; medical device; competitive inhibitor; pharmaceutical;	
KW	treatment; infection; septicemia; osteomyelitis; mastitis; endocarditis;	
KW	extracellular matrix; vascular graft; vascular stent; vaccine;	
KW	intravenous catheter; artificial heart valve; cardiac assist device;	
KW	antibacterial.	
XX		
OS	Staphylococcus aureus.	
XX		
PN	W0927109-A2.	
XX		
PD	03-JUN-1999.	
XX		
PF	25-NOV-1998; 98WO-US25246.	
XX		
PR	31-AUG-1998; 98US-0098427.	
XX	26-NOV-1997; 97US-0066815.	
XX		
PA	(EIDH/) EIDHIN D N.	
PA	(FORF-) FORFAS T/A BIORESEARCH IRELAND.	
PA	(FOST/) FOSTER T J.	
PA	(HOOK/) HOOK M A O.	
PA	(INHI-) INHIBITEX INC.	
PA	(JOSE/) JOSEFSSON E.	
PA	(PATT/) PATTI J M.	
PA	(PERK/) PERKINS S E.	
XX		
PI	Eidhlin DN, Foster TJ, Hook MAO, Josefsson E, Patti JM;	
PI	Perkins SE;	
XX		
XX	WPI: 1999-357844/30.	
DR	N-PSDB; AAX77591.	
XX		
PT	Staphylococcus aureus fibrinogen-binding proteins for treating	
PT	septicemia, osteomyelitis, mastitis or endocarditis	

XX PS Claim 2; Fig 5; 143pp; English.

XX CC This invention describes novel Staphylococcus aureus fibrinogen-binding proteins that bind both the alpha and beta fibrinogen chains. The proteins (and their encoding nucleic acids are ClfB, SdrC, SdrD and SdrE). Staphylococcus aureus is thought to utilize fibrinogen to adhere to medical devices. Binding proteins that bind both the alpha and beta fibrinogen chains (ClfB, SdrC, SdrD and SdrE) can therefore be used as competitive inhibitors to block this binding. Antibodies against ClfB, SdrC, SdrD and SdrE inhibit ClfB, SdrC, SdrD and SdrE mediated binding. The proteins of the invention can be used in a pharmaceutical composition for the treatment of Staphylococcus aureus infection e.g. septicemia, osteomyelitis, mastitis or endocarditis or to inhibit the binding of S. aureus to the extracellular matrix. The proteins or their fragments may be used to coat a medical device to reduce the S. aureus infection of an indwelling medical device, especially where the medical device is selected from the group consisting of vascular grafts, vascular stents, intravenous catheters, artificial heart valves, and cardiac assist devices. ClfB, SdrC, SdrD, SdrE, or an active fragment, subdomain or encoding gene may be used as a vaccine. The DS (aspartate serine) repeat region or a gene encoding it may be used as an identifying probe for the identification of genes and encoding proteins from Staphylococcus aureus (other than ClfA), S. hemolyticus, S. lugdenensis, and S. schleriferi. The proteins of the invention have antibacterial activity.

XX SQ Sequence 918 AA;

Query Match 11.5%; Score 283; DB 20; Length 918;

Best Local Similarity 21.6%; Pred. No. 1.4e-10;

Matches 80; Conservative 75; Mismatches 187; Indels 28; Gaps 6;

QY 50 PGGSTKSHPEPQTPKPSKSSAEATQTPEDTPNK-----SGAEAKTKDSSNKS 98

Db 559 PEPSPDPEPEP-TPDPEPSPDPEPSPDPPDPSDSDSGSDSDSGSDSDSDSDS 617

QY 99 GAERAKTKGTSKSGEATQTKDSTKSHSELQTPKDTGKSGAEATQTPEDSPNRSGAE 158

Db 618 DS 677

QY 159 KTKQDPSKSGSQAQTKDVPNKGADGOTPKDGSKSGAEADOTPKDVPNKGAEATQTPK 218

Db 678 DSESDS 737

QY 219 DGSNKGAEQGPIDGPKSGAEQTSKDPNKKVVPQPSRKDHSPINSPNKGAEATQTPK 278

Db 738 DS 797

QY 279 DTNQLADKGLSPHAFKTESGEETDLISPPQEVKSEPTEDVEPKAEEDDTPGPEGGP 338

Db 798 DSDSDSDS-----DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSRVTP 849

QY 339 PKEEKEKMGSGASSENREGTLDSTGSEKDDLYPNGSGNGSAESSHFFAYLYTAALIVAV 398

Db 850 PNEQKAPNPKGEVHNKYSK---QHKTDALPTGDKSENTNATLFGAMM--ALLGSL 904

QY 399 LYIAHNKRK 408

Db 905 LTF---RKRK 911

RESULT 11

ID AAW89801

XX AAW89801 standard; Protein; 936 AA.

AC AAW89801;

XX AAW89801;

DT 16-MAR-1999 (first entry)

XX Staphylococcus aureus protein SEQ ID #5249.

DE Computer readable medium; vaccine; S.aureus infection; immunodetection;

KW Cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

XX

KW skin infection; surgical wound infection; scalded skin syndrome;

XX toxic shock syndrome.

OS Staphylococcus aureus.

XX EP786519-A2.

PN 30-JUL-1997.

XX 07-JAN-1997; 97EP-0100117.

XX 05-JAN-1996; 96US-0009861.

PR (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

PI Rosen CA;

XX WPI; 1997-374922/35.

DR Polynucleotide(s) and proteins derived from Staphylococcus aureus

XX stored on computer readable medium and used in the production of

PT anti-S.aureus vaccines

XX Claim 23; Page 3255-3258; 3271pp; English.

PS

XX This sequence represents a Staphylococcus aureus protein sequence of the

CC invention. The DNA sequences encoding the S.aureus proteins are recorded

CC on a computer readable medium, preferably selected from a floppy or hard

CC disk, random access memory (RAM), read-only memory (ROM) or CD-ROM.

CC Homology searches using the S.aureus DNA sequences allows putative

CC functions to be assigned so that protein-encoding or regulatory regions

CC of commercial, therapeutic or industrial importance can be obtained.

CC Specifically, sequences which are likely to encode antigens have been

CC identified and these polypeptides can be used in a vaccine composition

CC against S.aureus infection. The polypeptides can also be used in a kit

CC for the immunodetection of S.aureus in a sample. S.aureus is implicated

CC in numerous human diseases, including cellulitis, eyelid infections, food

CC poisoning, osteomyelitis, skin and surgical wound infections, scalded

CC skin syndrome, toxic shock syndrome, etc. Organisms transformed with the

CC DNA sequences can be used for recombinant production of the polypeptides.

CC The new DNA sequences (and their fragments) are useful as primers or

CC probes for isolating homologues of any of the 5191 S.aureus DNA sequences

CC contained on the computer readable medium.

XX

XX Sequence 936 AA;

QY Query Match 11.4%; Score 281; DB 18; Length 936;

Db Best Local Similarity 22.9%; Pred. No. 1.9e-10;

Matches 75; Conservative 73; Mismatches 170; Indels 10; Gaps 3;

QY 37 AGNVTHPSLSQRPGGSTKSHPEPQTPKPSKSSAEATQTPEDTPNKGAEATQKSSN 96

Db 590 SGSDSTSDSGSDSADS-DSASDSDSADSASDSDSDSDSDSDSDSDSDSDSDSDS 648

QY 97 KSGAEAKTKGTSKSGSQAQTKDSTKSHSELQTPKDTGKSGAEATQTPEDSPNRSGA 156

Db 649 DS 708

QY 157 EAKTKDPSKSGSQAQTKDVPNKGADGOTPKDGSKSGAEADOTPKDVPNKGAEATQ 216

Db 709 DS 768

QY 217 PKDGSNKGSAEQQPIDGPKSGAEQTSKDPNKKVVPQPSRKDHSPINSPNKGAEAT 276

Db 769 DS 824

QY 277 KADTNQLADKGLSPHAFKTESGEETDLISPPQEVKSEPTEDVE-----PKEAEDDT 331

Db 825 DS 884

QY 332 GPEEGSPPKKEEKKMGSGASSENREGTL 359

XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
PS
PS
XX Example 4; SEQ ID NO: 28374; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
XX Sequence 617 AA;

Query Match 11.1%; Score 272.5; DB 22; Length 617;
Best Local Similarity 22.9%; Pred No 4.4e-10;
Matches 106; Conservative 66; Mismatches 185; Indels 105; Gaps 13;

QY 26 VKOEEAGVRPSAGNVSTHPSLSPRGSGTSHPEPOTPKDPS-----KSSAEAPQTPE 79
DB 141 VKSPEAKSPAKEEASPPKAESPEREEAKSPAEVKSPEAKSPAKEEAKSPAESPEK 200
QY 80 TPN-----KSGAEATKQDSNNKSGAEATQGSTSKSGSEAAQTDTSDTSKSHSETTP 133
DB 201 AKSPVKEEAASPAEKSPVKKEAASPAEVKSPEAKSPTTKEEAKSPAESPEAKEAKSP 260
QY 134 KDSGTGSGAEAOPTPE-----DSPNRSGAETKTOKDSPSKSGSEAQTTRKDVPNKSG 183
DB 261 EKASPVKAEEKSPERAKSPVKAEEKSPERAKSPVKKEEAKSPAESPEAKSPEKAK 320
QY 184 A-----DGQTEPKDSSKSGADQTPKDVPNKSGAEKQTPKDGSNKSGAEQ--GPID--- 233
DB 321 SPVKEEAKTPEKASPVKKEEAKSPAESPEAKSPEKAKTLTDVKSPEAKTPEAKERSPADRFP 380
QY 234 -----GPSKGAEQTSKDSPNKVVP-----BQSRKD 261
DB 381 KAKSPVKEEVKSPERAKSLKADAKAPEKEIPKKEEVKSPVKEEPQEVKVKBPPEKAAE 440
QY 262 HSKPISNP-----SDNKELPKADTNQ---LADGKLSPHAFKTESGEETDLI 305
DB 441 EEKAPTPTKTEKKDKKKEEAPKKEAPKVEEKKPEPAVEKPKESKYEAKEEAEDKKV 500
QY 306 SPOQEEKVS-EPTEDVEPKE-----AEDDDTGPEGSP-----KEEK-- 343
DB 501 PTEKBAPAKVEVEDAKPEKTEVAKEFDPAKEDDAKAPSPAEKPAEKAEAPKKTDEKAK 560
QY 344 ---ERMSSGASSENRECTLSDSTGSEKDDILYPNGSGNGSAES 382
DB 561 KPREEKPTEAKEDDDTKLSKEFSKPAEAKESSSTDQKDS 602

RESULT 15
AAM16458
ID AAM16458 standard; Protein: 617 AA.
XX
XX AAM16458;
XX
XX
XX 12-OCT-2001 (first entry)
XX
XX
DE Peptide #2892 encoded by probe for measuring cervical gene expression.
DE Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
KW Homo sapiens.
XX OS
XX WO200157278-A2.
XX PN
XX PD
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000670.


```
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX PS Claim 27; SEQ ID No 21284; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 617 AA;
Query Match 11.1%; Score 272.5; DB 22; Length 617;
Best Local Similarity 22.9%; Pred. No. 4.4e-10;
Matches 106; Conservative 66; Mismatches 185; Indels 105; Gaps 13;
QY 26 VKQEEAGVRRPSAGNVSTHPSLSQRPGGTSHPEQTPKDSPPS-----KSSAEQAQTPED 79
DB 141 VKSPKAKSPAEEAKSPPEAKSPEKEEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 200
QY 80 TPN-----KSGAEAKTQKDSNKGGAETKQSGTSGSGEAQTTKSTSKSHSELQTP 133
DB 201 AKSPVKEAKSPAEEAKSPVKEEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 260
QY 134 KDSTGKSGAEQTPPE-----DSPNRSGAEAKTQKDSKSGSEAQTTKDVPNKSG 183
DB 261 EKAKSPVRAEAKSPPEAKSPVKAEEAKSPEKAKSPVKEEAKSPEKAKSPVKEEAKSPEKAK 320
QY 184 A-----DGOTPKDGSKSGAEDQTPKDVPNKSGAEKQTPKDGSKSGAEQ--GPID---- 233
DB 321 SPVKEAKTPKAKSPVKEEAKSPEKAKSPEKAKTLDVKSPEAKTPAKEEAKSPADKPPE 380
QY 234 -----GPSKSGAEQTSKDSPNKVP-----EOPSRKD 261
DB 381 KAKSPVKEEVKSPKAKSPLKADAKAPEIKPEKEEVKSPVKEEAKPQEVKVPKPPKAE 440
QY 262 HSKPTISNP-----SNKELPKADTNQ-----LADKGLSPHAPKTESGEETDLI 305
DB 441 EEKAPATPKTEEKDSKKEAPKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEK 500
QY 306 SPPOEEVKSS-EPTEDVEPKE-----AEDDDTGPEEGSPP-----KEEK-- 343
DB 501 PPEKEAPAKVEVKEDAKPEKTEVAKKEPDDAKAKPEKPAEKKEAPEAKKDTKEEKAK 560
QY 344 ----EKMMSGASSENREGTLDSTGSEKDDLYPNGSGNGSAES 382
DB 561 KPKEKPKTEAKAKEDDKTLSKEPKPAEKAKESSTDKDS 602
```

Search completed: March 4, 2003, 14:58:31
Job time : 46.1499 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:52 ; Search time 21.5931 Seconds
(without alignments)
2137.006 Million cell updates/sec

Title: US-09-763-902B-1
Perfect score: 2459
Sequence: 1 MRFVVALVLLNVAAGAVPL.....YIPLEKDERHOWIVLLSQFL 480

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	677.5	27.6	380	2 S22415	membrane protein T
2	673	27.4	361	2 B56940	integral membrane
3	668	27.2	363	2 A56940	integral membrane
4	306	12.4	1087	1 QFM5H	neurofilament trip
5	286.5	11.7	989	2 D89852	fibrinogen-binding
6	284	11.5	854	2 S02003	neurofilament trip
7	284	11.5	933	2 S41539	fibrinogen-binding
8	283	11.5	1072	1 A37221	neurofilament trip
9	274.5	11.2	877	2 F90070	Clumping factor B
10	273	11.1	606	2 A43427	neurofilament trip
11	272.5	11.1	1020	1 QFH0H	neurofilament trip
12	259	10.5	1664	2 T18262	S-layer protein -
13	256.5	10.4	332	2 B43427	neurofilament prot
14	251.5	10.2	406	2 S38170	SRP40 protein - ye
15	251	10.2	990	2 I51618	nucleolar phosphop
16	242.5	9.9	1092	2 T30214	fibrinogen-binding
17	242	9.8	734	2 B42680	nucleolar cytoplas
18	238.5	9.7	813	2 S70795	vsaa protein precu
19	238.5	9.7	971	2 T19431	hypothetical prote
20	236.5	9.6	916	2 A27864	neurofilament trip
21	233.5	9.5	635	2 T09648	nucleolin homolog
22	232	9.4	699	2 T38073	nucleolar phosphop
23	228	9.3	3507	2 T34513	hypothetical prote
24	226.5	9.2	488	2 I46014	cyclin II - bovin
25	226.5	9.2	489	2 A45988	dentin matrix acid
26	226.5	9.2	1337	2 T30291	dextranase - Strept
27	226.5	9.2	1385	2 D89824	hypothetical prote
28	225.5	9.2	6642	2 T29757	protein UNC-89 - C
29	224	9.1	5327	2 T13564	microtubule-associ

hypothetical prote
hypothetical prote
FmbB protein jampo
hypothetical prote
hypothetical prote
neurofilament medi
hypothetical prote
glucan 1,4-alpha-g
NF-180 - sea lamp
hypothetical prote
TCOF1 protein - mo
junctional sarcopt
stable tubule only
fibrinogen-binding
surface antigen Tc
neurofilament prot

ALIGNMENTS

RESULT 1
S22415
membrane protein TGN38 long form precursor - rat
N:Alternate names: membrane protein TGN41; trans Golgi network-specific 38K protein;
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 08-Oct-1999
C:Accession: S22415; S36285; S11468
R:Reaves, B.; Wilde, A.; Banting, G.
submitted to the EMBL Data Library, February 1992
A:Reference number: S22415
A:Accession: S22415
A:Molecule type: mRNA
A:Residues: 1-380 <REA>
A:Cross-references: EMBL:X64600; NID:g57350; PIDN:CAA45884.1; PID:g57351
R:Reaves, B.; Wilde, A.; Banting, G.
Biochem. J. 283 313-316, 1992
A:Title: Identification, molecular characterization and immunolocalization of an isof
A:Reference number: S36285; MUID:92246851; PMID:1575675
A:Accession: S36285
A:Molecule type: mRNA
A:Residues: 321-380 <REA2>
A:Cross-references: EMBL:X64600
R:Luzio, J.P.; Brake, B.; Banting, G.; Howell, K.E.; Braghetta, P.; Stanley, K.K.
Biochem. J. 270 97-102, 1990
A:Title: Identification, sequencing and expression of an integral membrane protein of
A:Reference number: S11468; MUID:90372941; PMID:2204342
A:Accession: S11468
A:Molecule type: mRNA
A:Residues: 1-334, 'LKL', <LUJ2>
A:Cross-references: EMBL:X53565; NID:g57346; PIDN:CAA37637.1; PID:g57347
R:Wong, S.H.; Hong, W.
J. Biol. Chem. 268 22853-22862, 1993
A:Title: The SYXORL sequence in the cytoplasmic domain of TGN38 plays a major role in
A:Reference number: A48718; MUID:94043053; PMID:8226795
A:Contents: annotation
C:Keywords: glycoprotein; transmembrane protein
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-380/Product: membrane protein TGN38 long form #status predicted <MAT>
F:18-303/Domain: trans-Golgi network luminal #status predicted <LUM>
F:304-324/Domain: transmembrane #status predicted <TMN>
F:325-380/Domain: intracellular #status predicted <CYT>
F:348-353/Region: trans-Golgi network localization signal
F:25,26,140,147,297/Binding site: carbohydrate (Asn) #status predicted

Query Match 27.6% Score 677.5; DB 2; Length 380;
Best Local Similarity 39.18; Pred. No. 7.2e-25;
Matches 183; Conservative
QY 1 MRFVVALVLLNVAAGAVPLLATESVKQEEAGVRPSAGNVSTHPSLSORPGGSKSP-- 58
DB 1 MQFLVALLLLSVAVARAL-----PSASKPNNTSSENNPPI 35

[illegible]

Db 1094 PEEPI7DTSPDEPTPSDEPTPSDEPTPSDEPTPEEPI7DTSPDEPTPS 115

S38170 SRP40 protein - yeast (*Saccharomyces cerevisiae*)
N...allrppctc...arag...roctid...vrrp002...proteic...vrn412-
.....

.....

C; Species: *Saccharomyces cerevisiae*
C; Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C; Accession: S38170; S40645; S37702
R; Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo, J. submitted to the Protein Sequence Database, March 1994
A; Reference number: S38158
A; Accession: S38170
A; Molecule type: DNA
A; Residues: 1-406 <BAL>
A; Cross-references: EMBL:Z28317; NID:g486580; PID:g486581; MIPS:YKR092c
A; Experimental source: strain S288C
R; Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; JI Yeast 9, 1349-1354, 1993
A; Title: The complete sequence of a 15 820 bp segment of *Saccharomyces cerevisiae* chromosome
A; Reference number: S40644; MUID:94205265; PMID:8154186
A; Accession: S40645
A; Molecule type: DNA
A; Residues: 1-406 <BOU>
A; Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51946.1; PID:g450552
A; Experimental source: strain S288C
R; Lalo, D.; Carles, C.; Sentenac, A.; Thuriaux, P.
submitted to the EMBL Data Library, May 1993
A; Description: Interactions between three common subunits of yeast RNA polymerases I and II
A; Reference number: S37702
A; Accession: S37702
A; Molecule type: DNA
A; Residues: 1-399, 'N', 401-406 <LAL>
A; Cross-references: EMBL:L11275; NID:g295670; PID:g295671
C; Genetics:
A; Gene: SGP:SRP40
A; Cross-references: SGD:S0001800; MIPS:YKR092c
A; Map position: 11R

```

Query Match      10.2%; Score 251.5; DB 2; Length 406;
Best Local Similarity 23.0%; Pred. No. 3.7e-05;
Matches 82; Conservative 67; Mismatches 167; Indels 41; Gaps 7;

QY 18 VPLLATESVKOEAGVPRPSAGNVSTHPSLSQRGGSTKSHPEQTPKDPSPKSGSAEAQTP 77
Db 11 VPKL---SVKEKEIEEKSSSSSSSSSS-----SSSSSSSSSSSSSG 50

QY 78 EDTPNKSGBAEAKTQKDSNKSNGAEAKTQKSTSKSGEATTKDSTKSHSELTQPKDST 137
Db 51 ESSSSSSSSSSSSSSSDSDSDSDSSSSSSSSSSSS-----SSSDSFESSDESDDSS 103

QY 138 GKSGAEAQTPEDSPNRSGAEAKTQK---DSPSKSGSBAQTQKVPNKSGADGOTPKDGSS 194
Db 104 GSSSSSSSSSDSESESEDETKKRARESDNEDAKETTKAKTEPESSESSSSSSSSSS 163

QY 195 KSGAEDOTPKDVPNKSGAEAKTQPKDGSNGSGAEEOGQIDGPSKSGAEQTSKOSPKNVVP 254
Db 164 SSESSEGSDDSDSSSSSSSSSSSGESDESDDSDSSSSSSSSSSSSSDSDSDSSSSSSSS--D 220

QY 255 EQPSRKDHSPISNPSDNKELPKADTNQADKQKLSPHAFKTESGEETDLISPPQEEVKS 314
Db 221 SDSSSSSSSSSSDSDSDSDSDSDSDSDSGSDSSSS---DSSSDSESTSDSDSDSDSD 276

QY 315 SEPTEDVPEKEADDGTGPEGSPPKKEKMGSGASSENK-----EGTLDSDTGSEK 367
Db 277 SGSSSELETKEATADESKAEATPASSNESTPPASSSSSSSSANKLNIPAGTDRIKEGQR 333

```

RESULT 15
I51618
nucleolar phosphoprotein - African clawed frog
C/Species: Xenopus laevis (African clawed frog)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C/Accession: I51618; S57757
R/Cairns, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A/Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A/Reference number: I51618; MUID:96019267; PMID:7593294
A/Accession: I51618

```

A:Status: preliminary; translated from GB/EMBL/DDBJ.
A:Molecule type: mRNA
A:Residues: 1-990 <CAI>
A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921
C:Genetics:
A:Gene: xNoop180
C:Superfamily: nucleolus-cytoplasm shuttle.phosphoprotein
C:Keywords: phosphoprotein

Query Match      10.2%; Score 251; DB 2; Length 990;
Best Local Similarity 24.3%; Pred. No. 9e-05;
Matches 108; Conservative 63; Mismatches 181; Indels 92; Gaps

Qy 22 ATESVKQEEAGVRSAGN-----VSTHPSLSQRGGSTKSHPPQPP-----KDS 66
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 453 SSDSSDEETTKPAAKTTPAKSAATPTSKTPTNSKATPTSKTTPAKPGTPKTSAAKDS 512
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 67 PKSSAAQTPPE-DTPNKGAEAKTKDSSNKGAE--AKT----KRGSTSKSGEAQTTK 170
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 513 SSDSSDSSSDEKTPAKRAKTTTPAKPAKTTTPAKPAKTTTPAKPAKSTPGKQVPTKK 572
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 121 DSTS--KXSHLOTPKDSTCKSGAEQTPEDSPNR-----SGAEAKTKQKDSPSKSGS 170
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 573 ESSSDSDSDSSSEDEKXSSAKPAVKT-TPGKATSKPVVASKPVPAKASSDSDSDEE 631
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 171 EAQTTKDVPNKSGA-----DGTPTKDGSKSGAEQDT-PKQVPNKGAEKQTPKDGSN 222
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 632 TTKTTTKPLTKLSPAVKTLPPKKAESSDSDSDSEKTKTPAKPPAKSATPVNTKAPAQN 631
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 223 K-----SGAEQGPIDGGS-KSGAEQTSKDSPNKVPPEQPSRKDHSPINPSD 271
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 692 KASKASCSDSDSSEEGKSKQPTGKSPAAKATAPPKNVPVANKDPSSSSSDSGDD 751
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 272 NKELPK-----ADPNQADKGLSPHAFKTESGETDLISPPQEEVKSSEPTEDV----- 321
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 752 EKQPKQAAAAKDVQGAKAAPTPKKAASSSSD-----SSDEDVSKAKKTTAVSKSP 807
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 322 -----EPKAEDDDTGPGE-----GSPKKEEKMKMSGASSENREGT 358
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 808 VTPKAVPAAKKSSSSSSDSEDEKQGGKNTTTKIANSTPKAAAAECSESSSSDEGK 867
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 359 LSDTSGSEKDDLYPNGSGNGSAES 382
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 868 ANGTSGKKRRE-----STGNAFCEA 887
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |

Search completed: March 4, 2003, 15:04:25
Job time : 25.5931 secs

```

Search completed: March 4, 2003, 15:04:25
Job time : 25.5931 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	284	11.5	933	3	US-08-293-728-2	Sequence 2, Appl1
2	284	11.5	933	4	US-09-421-868-2	Sequence 2, Appl1
3	219	8.9	754	4	US-09-214-564A-2	Sequence 2, Appl1
4	208.5	8.5	571	4	US-09-134-001C-3865	Sequence 3665, Ap
5	208.5	8.5	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
6	205.5	8.4	1187	1	US-08-320-559-28	Sequence 28, Appl
7	205.5	8.4	1187	3	US-08-545-860D-28	Sequence 28, Appl
8	205.5	8.4	1187	5	PCT-US94-04496-28	Sequence 28, Appl
9	205.5	8.4	1210	1	US-08-320-559-26	Sequence 26, Appl
10	205.5	8.4	1210	3	US-08-545-860D-26	Sequence 26, Appl
11	205.5	8.4	1210	5	PCT-US94-04496-26	Sequence 26, Appl
12	204	8.3	1581	4	US-09-110-517-2	Sequence 2, Appl1
13	199	8.1	559	1	US-08-320-559-31	Sequence 31, Appl
14	199	8.1	559	3	US-08-545-860D-31	Sequence 31, Appl
15	199	8.1	559	5	PCT-US94-04496-31	Sequence 31, Appl
16	190.5	7.7	1060	4	US-08-911-393-2	Sequence 2, Appl1
17	188	7.6	976	4	US-08-894-997-50	Sequence 50, Appl
18	188	7.6	1848	4	US-08-296-791-6	Sequence 6, Appl1
19	188	7.6	1848	5	PCT-US95-10661A-6	Sequence 6, Appl1
20	184.5	7.5	1776	4	US-09-556-877-179	Sequence 179, App
21	184.5	7.5	1776	4	US-09-620-412C-179	Sequence 179, App
22	182.5	7.4	688	3	US-09-141-047-8	Sequence 8, Appl1
23	179.5	7.3	684	1	US-08-421-661-6	Sequence 6, Appl1
24	177	7.2	568	1	US-08-320-559-30	Sequence 30, Appl
25	177	7.2	568	3	US-08-545-860D-30	Sequence 30, Appl
26	177	7.2	568	5	PCT-US94-04496-30	Sequence 30, Appl
27	177	7.2	778	6	5198347-4	Patent No. 5198347

```

; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 754
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-564A-2

Query Match      8.9%; Score 219; DB 4; Length 754;
Best Local Similarity 24.1%; Pred. No. 9.1e-09;
Matches 90; Conservative. 66; Mismatches 162; Indels 56; Gaps 15;

QY 35 PS-AGNVSTHPSLSORPGGTTKSHPEPQTQKDS-----SKSGAEAOQTEDTNP 82
Db 395 PSRAHFESEHPTSKMNAQETATGMAFEEPIDEKKTEKTEGGGEPEGASENNAKESPE 454
QY 83 KSGAEAKTQKDSN---KSGAEAKTQKSTKSGSEAQTTKDTSKSHSELQTPKDSQTK 139
Db 455 KEAEEGCPKESEEGCPKGFEGSCQ-KESEEGNPNVRSSEEDSPKSKKTKLNDCDEE 513
QY 140 SGAEAOQTPEPSNRSKSGAEAKTQKDSKSGSEAQTTKDVPNKGADG---QTPKDGSSKS 196
Db 514 NGLAKESDNLNKESEEEVGPTKES-EEDDSKESEDEDCSEKQSEGGSEEFEEENGLEKD 572
QY 197 GAEDQTPKQVPNKGAEKQTPKDGSKNGAEBOGPIDGPSPKSGAEEOQTSKDPNKPWPEQ 256
Db 573 LDEEGSEKEL-HENVLDKLELNDSENSEFEDGGSEKVLDEGSEREFDEDSDEK-----E 627
QY 257 PSKDRSKSPISNPNKELPKADTNQADKGLSPHAKFTSG-----EETDLISPPQ 309
Db 628 EEDDIYEKVFDDSEDEKE-----DEYADEKLEAAADKAEAGDADEKLFEESD---DKE 679
QY 310 EVKVSSEPTEDYEPKEAEDDTG-----PEEGSPPK-----EEKEKMSGSSASSENRECTL 359
Db 680 DEDADGKEVEDADEKLFEEDDSNEKLFDEEDSSEKLFDDSDERGTGGFSGVE--EGPL 737
QY 360 SDST-----GSEKDD 369
Db 738 STGSSFILSSDDDD 751

RESULT 4
US-09-134-001C-3865
; Sequence 3865, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCI
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3865
; LENGTH: 571
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3865

Query Match      8.5%; Score 208.5; DB 4; Length 571;
Best Local Similarity 20.4%; Pred. No. 4e-08;
Matches 75; Conservative 64; Mismatches 186; Indels 43; Gaps 11;

QY 29 EEAGVPSAGNVSTHPSLSORPGGTTKSHPEPQTQKDSKSGSAEAOQTEDTNPKSGAEA 88
Db 163 EQNGGKPATSTTGQDAAQNSNGPGTSSVAPQ-----GNDASSQGHDTFQADAGTHTEGQA 218
QY 89 KTKQDSNKSNGAEAKTQKSTKSGSEAQTTKDTSKSHSELQTPKDSQTKSGAGAA----- 144

```

Db 219 VT-----TGEOTGKPAATKPEGASATAPAPNADENHAOPSOTGGTTATQAGQVK 271
QY 145 -----QTPESPNSRGAETKQDPSKSGSEATTKDVPNK--SGADG---QTPKDGSSK 195
Db 272 PSGDSTPTGPNKAD-DODTQIKPTSNQGTTA-TTNEIGNQKPSQGTGNTENTPNDGTQV 329
QY 196 SGAEQDTPKDVPNKSGAEKOTPKDGNKSGAEBOGPIDGPSKSG-----AEQTSKDSP 249
Db 330 KPNNDPAVGTPTNNG--BOTGPNATVPGQGNNEINGASKPGEVSPKPEENNPATPEP 387
QY 250 NKVVPQPSRKDHKSISNP-----SDNKELPKADTNQLADKGLSPHAFKTESGEE 301
Db 388 GTAPANTNQDQVKENTQATGTAGTDNNTQOQNTQONNQAQPSAPGTTDQAGAT 447
QY 302 TDLISPPQBEVYSS-EPTEDVEPEKAEDDDTGPESGPKPEKEKMGSGSASSNREGTLLS 360
Db 448 VRPGSAPNQDAEATPNPDQNTPTNGDQDOTNQSTQDDNDNQNTQOQNTKQNNQNAEQG 507
QY 361 DSGSEKD 368
Db 508 NTGGTDKD 515

RESULT 5
US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4463

Query Match 8.5%; Score 208.5; DB 4; Length 2137;
Best Local Similarity 22.8%; Pred. No. 2.2e-07;
Matches 93; Conservative 77; Mismatches 197; Indels 41; Gaps 11;

QY 10 LNVAAGAVPLLATESYKQBEAGVRPSAGNVSTHPSLSQRPGGSTKSHPEPQT----- 62
Db 863 LSESASTSDSASASARKSESTKSTSLSESTSTSVSDSASVSTSTSVSGSTS 922
QY 63 -----PKDPS-KSSAEQTP-----DTPNKGAEAKTQDSSNKGAEAKTQ 105
Db 923 TSISDSTSTSDSASIKASASASTSKLSESVSTSTSDASTSTSVSDSN-SASTSLSK 981
QY 106 KGSTSKSGEAQTTKSTKSHSLOTPKDTGKSGAEAOPTDPSPNRSGAEAKTQKDSP 165
Db 982 STSVSDSTSTSDASTSTSESDSASTSLSESTSTSVSDSTSTSDSASMSASE 1041
QY 166 SKSGSEAKTKDVPNK--SGADGOTPKDGSKSGAEAOPTKDVPNKSGAEAKTQPDGSKN 223
Db 1042 SESNKSSTLSSESTSTSLSGSTASTSDASTSTSESES--DSTSTLSSESTSTSL 1095
QY 224 SGAEQPIDGPSKSGAEQTKSPKVPYEQPSRKDHSPKIPSPDNKELPKADTNOL 283
Db 1096 SGSTASTSDASTSTSESDSTSLSESTSTSVSDSTASTSESTSTSESN 1155
QY 284 AD--KGKLSPHAFKTESGEEDTLSPPOEYKSGEPTEDVEPKAE--DDDTG---PREG 336
Db 1156 STSLSGSLSTISDSTSTSTSDASTSTSESDSTSTSLSESTSTSLSDSTSTSTSESA 1215

QY 337 SPPKEKMKSGAS--SENREGLTSDTSGSEKDDLYPNNGSGMSAES 383
Db 1216 STSTSDSTSESTSLSESTSTSVSDSTASTSDASTSTSVSDSESA 1263

RESULT 6
US-08-320-559-28
; Sequence 28, Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
; TITLE OF INVENTION: All-1 Region
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320,559
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-320-559-28

Query Match 8.4%; Score 205.5; DB 1; Length 1187;
Best Local Similarity 23.4%; Pred. No. 1.7e-07;
Matches 114; Conservative 52; Mismatches 188; Indels 133; Gaps 21;

QY 19 PLLATESYKQBEAGVRPSAGNVSTHPSL-----SQRPGGSTKSHPEPQPKD----- 65
Db 347 PLTAIHPTSAEPKSEFPPTKDSQHVSVTQNKQYDTSSTKTHSNSQGTSSMLEDLQL 406
QY 66 SPKSSAEAOPTDTPNKGAEAKTQK-----DSNKGAEAKTQKSTKSGSEAOPTTK 120
Db 407 SDSESDSEQTPKPPSPSSAPPSQSLPEPVASAHSSAESESTSDSDSSDSESS- 465

QY 121 DSTSKSHSELOTPKDS-----TGKSGAEAOPTPE-----DSPNRSGAEAKTQK 162
Db 466 SSDESENEPLETAPPEPTTNKWQDLNLTQVSOAPAPPEGRSTPEPRRHPESKGS 525
QY 163 DSPSKSGSEAOITKDVNPKSG-----ADGOTPKDGS-----SKSGAEAOPTPK-----DVPN 208
Db 526 D--SATSQHSESKDPPKSSKAPRAPPEPHGRSCOKSPAQQEPPOQTGTPKPK 583
QY 209 KSGAEKOTPKDGNKSGAEQPIDGPKSGAEQTSKD-----SPNKV 252
Db 584 K--PVKASARAGRTSLOGREP--GLLPYGSRDQTSKORPKVKTGRPRAAASNEPKPA 639
QY 253 VPEQPSRKDHSPISNPSDNKELPKADTNQADKGLSPHAFK-----TE-----SCE 300
Db 640 VPPSEKSKHKSLLPAPSKALSGPEPAKDNVEDR---TPEHFALVPLTESQGPPIHSGSGS 696
QY 301 ET-----DLISPPQEVKSSSEPTEDVEP-----323
Db 697 RTSGCRQAVVQEDSRKDRPLPLRDTKLLSLRTPPPQSLMVKITLILSRIPQPGK 756
QY 324 --KEAEDDDTGPEEGSPKKEKMGSSASS---ENREGTLDSTGSEKDDLYPN--GSGN 377
Db 757 GSRQRAEDKQPPAGK--KHSSEKSSDSSSLAKRKGAEARDCDNKKIRLEKEIKSOS 814
QY 378 GSAESSH 384
Db 815 SSSSSSH 821

RESULT 7
US-08-545-860D-28
; Sequence 28, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaani, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545.860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093
FILING DATE: 11-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1262
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-545-860D-28

Query Match 8.4%; Score 205.5; DB 3; Length 1187;
Best Local Similarity 23.4%; Pred. No. 1.7e-07;
Matches 114; Conservative 52; Mismatches 188; Indels 133; Gaps 21;

QY 19 PLLATESVKQEEAGVRPSAGNVSTHPSL---SQRPGGSKSHPEPOTPKD-----65
Db 347 PLTAHTPTSTAPSKFPPTKDSQHVSVTQKQDTSKTHSNSSQOQTSSMLEDLQL 406
QY 66 SPSKSAEAOITPEDTPNKSGAEAKTQK-----DSSNKSGAEAKTQKSGSEAOITK 120
Db 407 SDESDSDSEOTPEKPPSSAPPSPQSLPEPVASAHSSAESSESTSDSSDSESESS- 465
QY 121 DSTSKSHSELOTPKDS-----TGKSGAEAOPTPE-----DSPNRSGAEAKTQK 162
Db 466 SDSENEPLETAPPEPTTNKWQDLNLTQVSOAPAPPEGRSTPEPRRHPESKGS- 525
QY 163 DSPSKSGSEAOITKDVNPKSG-----ADGOTPKDGS---SKSGAEAOPTPK-----DVPN 208
Db 526 D--SATSQHSESKDPPKSSKAPRAPPEPHGRSCOKSPAQQEPPOQTGTPKPK 583
QY 209 KSGAEKOTPKDGNKSGAEQPIDGPKSGAEQTSKD-----SPNKV 252
Db 584 K--PVKASARAGRTSLOGREP--GLLPYGSRDQTSKORPKVKTGRPRAAASNEPKPA 639
QY 253 VPEQPSRKDHSPISNPSDNKELPKADTNQADKGLSPHAFK-----TE-----SCE 300
Db 640 VPPSEKSKHKSLLPAPSKALSGPEPAKDNVEDR---TPEHFALVPLTESQGPPIHSGSGS 696
QY 301 ET-----DLISPPQEVKSSSEPTEDVEP-----323
Db 697 RTSGCRQAVVQEDSRKDRPLPLRDTKLLSLRTPPPQSLMVKITLILSRIPQPGK 756
QY 324 --KEAEDDDTGPEEGSPKKEKMGSSASS---ENREGTLDSTGSEKDDLYPN--GSGN 377
Db 757 GSRQRAEDKQPPAGK--KHSSEKSSDSSSLAKRKGAEARDCDNKKIRLEKEIKSOS 814
QY 378 GSAESSH 384
Db 815 SSSSSSH 821

RESULT 8
PCT-US94-04496-28
; Sequence 28, Application PCT/US9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaani, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1

NUMBER OF SEQUENCES: 86.
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESSEE: Norris
STREET: One Liberty place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DeLuca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-04496-28

Query Match 8.4%; Score 205.5; DB 5; Length 1187;
Best Local Similarity 23.4%; Pred. No. 1.7e-07;
Matches 114; Conservative 52; Mismatches 188; Indels 133; Gaps 21;

QY 19 PLLATESVKQEEAGVRPSAGNVSTHPSL-----SQRGGSTKSHPEQTPKD----- 65
Db 347 PLTAHTPSTAEPFKPFPTKDSQHVSVTQKQVDTSKTHSNQOGTSSMLEDLQL 406
QY 66 SPKSGAEATPDTPNKSGAEAKTK-----DSSNKSGAEAKTKGTSKSGSEATTK 120
Db 407 SDESDSDSEQTEKPPSSAPPAPQSLPEPVASAHSSAESESTSDSDSSDSESS- 465
QY 121 DSTKSHSELQTPKDS-----TGKSGAEATPE-----DSPNRSGAEAKTK 162
Db 466 SSDSENEPLETPAPEPEPTTNKWLDMTKVSOAPAPPEGPSTPRRHPESKSGS 525
QY 163 DSPSKSGSEATTKDVPNKG-----ADGTPKDG-----SKSGAEDQTPK-----DVPN 208
Db 526 D--SATSQEHSESKDPPPKSSKAPRAPPEAPHPGRKRSQKSPAQQEPQRTVGTQPK 583
QY 209 KSGAEQTPKDGSKSGAEQGPIDGPKSGAEQTSKD-----SPNKV 252
Db 584 K--PVKASARAGSRTSLOGREP--GLLPVGSRDQTSKQPKVTKGRPRAASNEPKPA 639
QY 253 VPQPKRKHSHKIPSNPDNKLPLKADTNOLADKGLSPHAFK-----TE-----SGE 300
Db 640 VPPSEKKKHKSLPAPSKALSQPEPAKDNVEDR--TPHFALVPLTESQGGPPHSGSGS 696
QY 301 ET-----DLISPPQEEVKSEPTEDVEP----- 323
Db 697 RTSGCQAVVQEDSRKDRPLRLDTLTKLSPLRDTPPQSLMVKITLDLLSRIPQPGK 756
QY 324 --KEAEDDDTGPEGSPKPEEKMGSGASS--ENREGTLDSTGSEKDDLYPN--GSGN 377
Db 757 GSKRKAEDQPPAGK--KHSSEKSSDSSKLAKRRKGAERDCDNKKIRLEIKSQS 814
QY 378 GSAESSH 384
Db 815 SSSSSSH 821

RESULT 9
US-08-320-559-26
Sequence 26; Application US/08320559
Patent No. 5633135
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaani, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
TITLE OF INVENTION: All-1 Region
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135rls.
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/320,559
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/062,443
FILING DATE: 14 MAY 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/971,094
FILING DATE: 30-OCT-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/888,830
FILING DATE: 27-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/805,093
FILING DATE: 11-DEC-91
ATTORNEY/AGENT INFORMATION:
NAME: DeLuca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-0855
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1210 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-320-559-26

Query Match 8.4%; Score 205.5; DB 1; Length 1210;
Best Local Similarity 23.4%; Pred. No. 1.8e-07;
Matches 114; Conservative 52; Mismatches 188; Indels 133; Gaps 21;

QY 19 PLLATESVKQEEAGVRPSAGNVSTHPSL-----SQRGGSTKSHPEQTPKD----- 65
Db 370 PLTAHTPSTAEPFKPFPTKDSQHVSVTQKQVDTSKTHSNQOGTSSMLEDLQL 429
QY 66 SPKSGAEATPDTPNKSGAEAKTK-----DSSNKSGAEAKTKGTSKSGSEATTK 120
Db 430 SDESDSDSEQTEKPPSSAPPAPQSLPEPVASAHSSAESESTSDSDSSDSESS- 488
QY 121 DSTKSHSELQTPKDS-----TGKSGAEATPE-----DSPNRSGAEAKTK 162
Db 489 SSDSENEPLETPAPEPEPTTNKWLDMTKVSOAPAPPEGPSTPRRHPESKSGS 548
QY 163 DSPSKSGSEATTKDVPNKG-----ADGTPKDG-----SKSGAEDQTPK-----DVPN 208

Db 549 D--SATSQHSSEKPPPKSSSKAPRAPEAPHKRCQKSPAQOEPQRTVGTGKPK 606
QY 209 KSGAEKQTPKDGSKNGAEQPIDGPKSGAEQTSKD-----SPNKV 252
Db 607 K--PVKASARAGSRISLOGEREP--GLLPYGRDQTSKDKPKVTKGRPRAAASNEPKPA 662
QY 253 VPEQSRKDHSPINSPDNKELPRADTNQADKGLSPHAFK-----TE-----SGE 300
Db 663 VPPSEKKKKHSSLPAPSKALSGPEAKDNVEDR---TPEHFALVPLTESQGPPIHSGSGS 719
QY 301 ET-----DLISPPQEVKSSEPTEDVEP----- 323
Db 720 RTSGCROAVVQEDSRKDRPLPLRDTKLLSPRTPPQSLMWKITLILSRIPQPPGK 779
QY 324 --KEAEDDDTGPEGSPKKEKMGSGASS---ENREGTILSDTSGSKDDLYPN--GSGN 377
Db 780 GSRQKAEQKOPPACK--KHSSEKSSDSSSKLAKRKGAEARDCDKNKIRLEKEIKSQS 837
QY 378 GSAESSH 384
Db 838 SSSSSSH 844

RESULT 10
US-08-545-860D-26
; Sequence 26, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaani, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESS: No. 6040140ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,094
; FILING DATE: 30-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,839
; FILING DATE: 27-MAY-1992

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,093
; FILING DATE: 11-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1210 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-545-860D-26

Query Match 8.4%; Score 205.5; DB 3; Length 1210;
Best Local Similarity 23.4%; Pred. No. 1.8e-07;
Matches 114; Conservative 52; Mismatches 188; Indels 133; Gaps 21;

QY 19 PLLATESYKQBEAGVRPSAGNVSTHPSL-----SORPGGSKTSKHPPEQPKD----- 65
Db 370 PLTAIHTPSTAEPKPEPTKDSQHVSVVTONQKQYDTSKTHSNSQOQTSMLDDQL 429
QY 66 SPSKSSAQAQPTEDTPNKSGAEAKTQK-----DSNKGAEAKTQKSTKSGSQAQTK 120
Db 430 SDESDSDSEOTPEKPPSSAPPQSLPEPVASAHSSAASESTSDSDSSDSESS- 488
QY 121 DSTKSHSELQTPKDS-----TKSGAEAKTPE---DSPNRSGAETQK 162
Db 489 SSDSENEPLETPEPEPPPTTNKQLDNWLTKVQAPPEGPRSTPEPRRHPSKSS 548
QY 163 DSPKSGSEATTKDVPNKSG-----ADGQTPKDGK---SKSGAEDQTPK-----DVPN 208
Db 549 D--SATSQHSSEKDPKPSKSSKAPRAPEAPHKRCQKSPAQOEPQRTVGTGKPK 606
QY 209 KSGAEKQTPKDGSKNGAEQPIDGPKSGAEQTSKD-----SPNKV 252
Db 607 K--PVKASARAGSRISLOGEREP--GLLPYGRDQTSKDKPKVTKGRPRAAASNEPKPA 662
QY 253 VPEQSRKDHSPINSPDNKELPRADTNQADKGLSPHAFK-----TE-----SGE 300
Db 663 VPPSEKKKKHSSLPAPSKALSGPEAKDNVEDR---TPEHFALVPLTESQGPPIHSGSGS 719
QY 301 ET-----DLISPPQEVKSSEPTEDVEP----- 323
Db 720 RTSGCROAVVQEDSRKDRPLPLRDTKLLSPRTPPQSLMWKITLILSRIPQPPGK 779
QY 324 --KEAEDDDTGPEGSPKKEKMGSGASS---ENREGTILSDTSGSKDDLYPN--GSGN 377
Db 780 GSRQKAEQKOPPACK--KHSSEKSSDSSSKLAKRKGAEARDCDKNKIRLEKEIKSQS 837
QY 378 GSAESSH 384
Db 838 SSSSSSH 844

RESULT 11
PCT-US94-04496-26
; Sequence 26, Application PCT/US9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaani, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESS: Norris
; STREET: One Liberty Place, 46th floor

CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/04496
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Deluca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TJU-1242
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1210 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-04496-26

Query Match 8.4%; Score 205.5; DB 5; Length 1210;
 Best Local Similarity 23.4%; Pred. No. 1.8e-07;
 Matches 114; Conservative 52; Mismatches 186; Indels 133; Gaps 21;
 QY 19 PLLATESVKQEEAGVRPSAGNYSTHPL-----SQRPGGSKTSHPEPQTPKD----- 65
 Db 370 PLTAHTPSTAEPKFPFPTKDSQHVSVTQKQYDTSSKTHSNQOGTSSMLEDDLQ 429
 QY 66 SPKSKSAEAQTEDTPNKGAEAKTQK-----DSSNKGAEAKTQKSTKSGSEATQTK 120
 Db 430 SDESDSDSQTPTEKPPSSAPPSPQSLPEPVASAHSSAESESTSDSDSDSESS- 488
 QY 121 DSTSKSHSLQTPKDS-----TGKSGAEATPE-----DSPNRSGAEAKTQK 162
 Db 489 SDDSENEPLETPAPEPEPTTNKNQWLNLTQVSPAPPEPSTPEPRRHPESKSGS 548
 QY 163 DSPSKSGSEATQTKDVPNKG-----ADQTPKDGs-----SKSGAEQTPK-----DVFN 208
 Db 549 D--SATSQEHSSEKDPKPKSSKAPRAPPEAPHGKRSCOKSPAQQEPQRTQVGTQPK 606
 QY 209 KSGAEQTPKDSNKGSGAEQGPIDGPKSGAEQTSKD-----SPNKV 252
 Db 607 K--PVKASARAGSRTSLOGEREP--GLLPYGSRDQTSKDKPKYKTKGRPRAAASNEPKPA 662
 QY 253 VPEQPSRKDPSKPSNPNKELPKADTNQADKGLSPHAFK-----TE-----SGE 300
 Db 663 VPPSEKKKKHKSLLPAPSKALSGPEPAKQNVDR---TPEHFALVLTESQGPSPHSGS 719
 QY 301 ET-----DLISPPQEEVKSEPTEDVEP----- 323
 Db 720 RTSGCRQAVVQEDSKRDLPLPLRTKLLSPLDTPPQSLMVKITLDDLRIPOPGK 779
 QY 324 --KEAEDDDTGPPEGPPKPEEKSGSASS-----ENRGTLSSTSEKDDLYPN--GSGN 377
 Db 780 GSRQKAEDKQPPAGK--KHSSEKSSDSSSKLAKRKGAEKRDCKNKKIRLEKEIKSQS 837
 QY 378 GSAESSH 384
 Db 838 SSSSSSH 844

RESULT 12
 US-09-110-517-2
 ; Sequence 2, Application US/09110517A
 ; Patent No. 6248520

GENERAL INFORMATION:
 ; APPLICANT: Roeder, Robert G
 ; APPLICANT: Fondell, Joseph D
 ; APPLICANT: Yuan, Chao X
 ; APPLICANT: Ito, Mitsuhiro
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NUCLEAR HORMONE
 ; TITLE OF INVENTION: RECEPTOR COACTIVATORS AND USES THEREOF
 ; FILE REFERENCE: 600-1-224
 ; CURRENT APPLICATION NUMBER: US/09/110,517A
 ; CURRENT FILING DATE: 1998-07-06
 ; NUMBER OF SEQ ID NOS: 51
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 1581
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-110-517-2
 Query Match 8.3%; Score 204; DB 4; Length 1581;
 Best Local Similarity 20.8%; Pred. No. 3.2e-07;
 Matches 95; Conservative 56; Mismatches 197; Indels 108; Gaps 15;
 QY 19 PLLATESVKQEEAGVRPSAGNYSTH-----PSLSQRPGGSKTSHPEPQTPKDSPSKSSAE 73
 Db 956 PLLTTGDLGKEKTKQKRVKEGNGTSTLSGLDSDKPKGRSRTPTSDNGSKDKPKR--- 1012
 QY 74 AQTPEPTPNKGAEAKTQKDSNKGAEAKTQKSTKSGSEAO----- 117
 Db 1013 -----KKADTEGKSPSHSSNRPFTPTSTGGSKSPGSAGRSQTPPGVATPIPKIT 1064
 QY 118 -----TTKDSKSHSELOTPKDTGKSGAEAOPTEDSPNRSAGEAK--TKDSDPSKSGS 170
 Db 1065 IQIPKGTVMVGPXPSHSQ--YTSGGVSSSGSKSHSHSSSSSSASTSKMKSSKSGSS 1123
 QY 171 EAQTTKDVPNKSGADQTPKDGSSKSGAEDQTPKDVPNKSGAEAKOTPKDGNKSGAEQ 230
 Db 1124 SKLSSMYSSOGSSGSSQSKNSQSGK-----PGSSPTKHGLSSGSSSTKMKPG 1176
 QY 231 -----PIDGPSK-----GAEQTSKDSNKNVPEQPSRKDHSKPIPNPNKDEL 275
 Db 1177 KPSSLNPNLSKPNISPSHSRPPGGSKLA--SPMKPVPTPPSSKAKSPISSGSGSHM 1234
 QY 276 PRADTNQADKGLSPHAFKTESG-----EETDLISPPQEE---VKSEPTEDVEPK 324
 Db 1235 -----SGTSSSGMKSSGSLGSSLSQKT-----PPSSNSCTASSSSFSFGSSSM 1280
 QY 325 EAEDDDTGPPEGPPKPEEKSGSASENREGTLDSTGSEKDDLYPNNGSGNGSAESSH 384
 Db 1281 SSSQNGHSGSKGKSPSRNKKPSLTAVIDKLKHGVVTSGPGE--DPLDQMGVSTNSSSH 1338
 QY 385 FFAYLVTAAILVAVLYIAHNKRKIIAFVLEGRSK 420
 Db 1339 PMS-----SKHNSGGEFGQKREK 1357
 RESULT 13
 US-08-320-559-31
 ; Sequence 31, Application US/08320559
 ; Patent No. 5631135
 ; GENERAL INFORMATION:
 ; APPLICANT: Croce, Carlo
 ; APPLICANT: Canaan, Eli
 ; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
 ; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
 ; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
 ; TITLE OF INVENTION: All-1 Region
 ; NUMBER OF SEQUENCES: 44
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
 ; STREET: One Liberty Place - 46th Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA

```

;
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320,559
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TUJ-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 559 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-320-559-31

Query Match 8.1%; Score 199; DB 1; Length 559;
Best Local Similarity 24.5%; Pred. No. 2.1e-07;
Matches 96; Conservative 59; Mismatches 147; Indels 90; Gaps 18;

QY 18 VPLLATESVKQEEAGVRPSAG--NVSTHPSLSQRPGGTSKSHPEQTPKDSKSSAAQ 75
DB 164 IPLSAFDPKK---TKPSHGSKDANKESSKTSKPHKVTKEHRE-RPRKDSKSSSKREL 218

QY 76 TPEDTPNKGAEAKTKDSSNK--SGAEAKTKGTSKSGSE-----AQTTKDSTSKSH 127
DB 219 ERE-----QAKSKTSYTKLGGRLPKKEKAPPPKAAKPKMALKETKLESTSPN 270

QY 128 SELQTPKD---STGKSGAEAOPTPDSNRSAGAEAKTKDPSKSGSEAOQTTKDVPNKSGA 184
DB 271 GPPPPPPPPPRASSKRPATADSPKFS-----AKQKSSSK-GSRS-----A 311

QY 185 DQTPKDGSSKSGAEDQTPKDVNKGAE---KQTPKDGSKNKGAEBOGPID----- 233
DB 312 PETSPTSSSSSFSKPKAKDKSSRGEKVAESEPRAKKALEYEENSSEDEASFSES 371

QY 234 ---GPSKGAEBOTKSDSPKVVPOPKRDKHSKPSIPNSDNKLPKADTNLQADKGLS 290
DB 372 AOSFSNSSSSSDSSDS-----DFEPONHSQGLRSMVEDLQSEEDD-----DSSSGE 423

QY 291 PHAFTEGSEETDLT---SPPOEEYKSSSEPTEDVEPKAEADDDTGPESGPPKKEKMS 347
DB 424 EAAGKNPCGRDRLSFSDESNSADSLPSREPPP-----PQRPPPP---NSKVS 471

QY 348 GSASSEN-----REGTSLDSTGSEKDDLY 371
DB 472 GRRSPESCSPKELIKKGTYDKAYTDDELVELH 503

RESULT 14
US-08-545-860D-31

; Sequence 31, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaani, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESS: No. 6040140ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,839
; FILING DATE: 27-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,093
; FILING DATE: 11-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TUJ-1262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 559 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-545-860D-31

Query Match 8.1%; Score 199; DB 3; Length 559;
Best Local Similarity 24.5%; Pred. No. 2.1e-07;
Matches 96; Conservative 59; Mismatches 147; Indels 90; Gaps 18;

QY 18 VPLLATESVKQEEAGVRPSAG--NVSTHPSLSQRPGGTSKSHPEQTPKDSKSSAAQ 75
DB 164 IPLSAFDPKK---TKPSHGSKDANKESSKTSKPHKVTKEHRE-RPRKDSKSSSKREL 218
```


Query Match 8.1%; Score 199; DB 5; Length 559;
Best Local Similarity 24.5%; Pred. No. 2.1e-07;
Matches 96; Conservative 59; Mismatches 147; Indels 90; Gaps 18;

Search completed: March 4, 2003, 15:05:59
Job time : 23.6405 secs

Search completed: March 4, 2003, 15:05:59
Job time : 23.6405 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:58:57 ; Search time 11.4899 Seconds

(without alignments)
1761.695 Million cell updates/sec

Title: US-09-763-902B-1

Perfect score: 2459

Sequence: 1 MRFVAVLLNVAAGAVPL.....YIPLEKDERHQWIVLLSFL 480

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications-AA:

- 1: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubaa/FCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubaa/FCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	302	12.3	1021	10	US-09-815-242-5471
2	302	12.3	1021	10	US-09-815-242-12544
3	272.5	11.1	617	10	US-09-864-761-36182
4	260.5	10.6	265	12	US-10-073-256-78
5	216	8.8	2344	10	US-09-815-242-12713
6	214.5	8.7	1367	10	US-09-801-368-108
7	212	8.6	1349	10	US-09-815-242-5898
8	212	8.6	1349	10	US-09-815-242-13137
9	211.5	8.6	2478	10	US-09-815-242-5816
10	211.5	8.6	2478	10	US-09-815-242-12967
11	210.5	8.6	1163	9	US-09-932-257A-3
12	210	8.5	932	10	US-09-815-242-5578
13	210	8.5	932	10	US-09-815-242-12438
14	202.5	8.2	724	9	US-10-108-605-307
15	201	8.2	2368	10	US-09-815-242-5635
16	201	8.2	2368	10	US-09-815-242-12389
17	198.5	8.1	1113	10	US-09-815-242-12769
18	198.5	8.1	1113	10	US-09-815-242-5836
19	198.5	8.1	1770	10	US-09-841-132-444

20	196.5	8.0	596	9	US-10-063-547-100	Sequence 100, App
21	196.5	8.0	596	9	US-10-174-590-310	Sequence 310, App
22	196.5	8.0	596	9	US-10-176-758-310	Sequence 310, App
23	196.5	8.0	596	9	US-10-063-616-100	Sequence 100, App
24	196.5	8.0	596	9	US-10-175-737-310	Sequence 310, App
25	196.5	8.0	596	9	US-10-063-502-100	Sequence 100, App
26	196.5	8.0	596	9	US-10-173-706-310	Sequence 310, App
27	196.5	8.0	596	9	US-10-175-738-310	Sequence 310, App
28	196.5	8.0	596	9	US-10-175-752-310	Sequence 310, App
29	196.5	8.0	596	9	US-10-176-482-310	Sequence 310, App
30	196.5	8.0	596	9	US-10-176-757-310	Sequence 310, App
31	196.5	8.0	596	9	US-10-176-913-310	Sequence 310, App
32	196.5	8.0	596	9	US-10-180-552-310	Sequence 310, App
33	196.5	8.0	596	9	US-10-180-557-310	Sequence 310, App
34	196.5	8.0	596	9	US-10-173-700-310	Sequence 310, App
35	196.5	8.0	596	9	US-10-174-572-310	Sequence 310, App
36	196.5	8.0	596	9	US-10-174-579-310	Sequence 310, App
37	196.5	8.0	596	9	US-10-174-582-310	Sequence 310, App
38	196.5	8.0	596	9	US-10-174-588-310	Sequence 310, App
39	196.5	8.0	596	9	US-10-175-739-310	Sequence 310, App
40	196.5	8.0	596	9	US-10-175-740-310	Sequence 310, App
41	196.5	8.0	596	9	US-10-175-743-310	Sequence 310, App
42	196.5	8.0	596	9	US-10-176-488-310	Sequence 310, App
43	196.5	8.0	596	9	US-10-176-492-310	Sequence 310, App
44	196.5	8.0	596	9	US-10-176-747-310	Sequence 310, App
45	196.5	8.0	596	9	US-10-176-750-310	Sequence 310, App

ALIGNMENTS

RESULT 1
US-09-815-242-5471
; Sequence 5471, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5471
; LENGTH: 1021
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5471

Query Match 12.3%; Score 302; DB 10; Length 1021;
Best Local Similarity 22.3%; Pred. No. 1.7e-09;

Matches 84; Conservative 81; Mismatches 188; Indels 24; Gaps 4;

QY 22 ATESVKQEEAGVRPSAGNVSTHPSLSORPGS-----TKSHPEPOTPKDSPSKSAEA 74
 Db 606 ASDSDSTSDSDSASDSDSASDSDSASDSDSASDSDSASDSDSASDSDSDSDSDSDSDS 665

QY 75 QTPEPTPKSGAEAKTQKDSNKGAEAKTQKSTKSGSEAOITKDTKSHSELOTPK 134
 Db 666 DS 725

QY 135 DSTGKGAEAOITPEDSPNRSGAEAKTQKDSNKGAEAKTQKSTKSGSEAOITKDVPNKSGADGOTPKDGS 194
 Db 726 DS 785

QY 195 KSGAEAOITPKDVPNKSGAEAKTQKDSNKGAEAKTQKSTKSGSEAOITKDVPNKSGADGOTPKDGS 254
 Db 786 DS 845

QY 255 EQPSRKDHKPIPSNDKELPKADTNQADKGLSPHAFKTESGETDLISPPQEEVKS 314
 Db 846 ESDS 896

QY 315 SEPTEDVEPKAEADDDTGPEEGSPKKEEKMGSGSAS-----SENREGLSDS-----TGSE 366
 Db 897 SDDS 956

QY 367 KDDLYPNGSGNGSAESS 383
 Db 957 NNVPNPSPKNGTASN 973

RESULT 2

US-09-815-242-12544
 ; Sequence 12544, Application US/09815242
 ; Patent No. US2002006159A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA-011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12544
 ; LENGTH: 1021
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-12544

Query Match 12.3%; Score 302; DB 10; Length 1021;
 Best Local Similarity 22.3%; Pred. No. 1.7e-09;

Matches 84; Conservative 81; Mismatches 188; Indels 24; Gaps 4;

QY 22 ATESVKQEEAGVRPSAGNVSTHPSLSORPGS-----TKSHPEPOTPKDSPSKSAEA 74
 Db 606 ASDSDSTSDSDSASDSDSASDSDSASDSDSASDSDSASDSDSASDSDSDSDSDSDSDS 665

QY 75 QTPEPTPKSGAEAKTQKDSNKGAEAKTQKSTKSGSEAOITKDTKSHSELOTPK 134
 Db 666 DS 725

QY 135 DSTGKGAEAOITPEDSPNRSGAEAKTQKDSNKGAEAKTQKSTKSGSEAOITKDVPNKSGADGOTPKDGS 194
 Db 726 DS 785

QY 195 KSGAEAOITPKDVPNKSGAEAKTQKDSNKGAEAKTQKSTKSGSEAOITKDVPNKSGADGOTPKDGS 254
 Db 786 DS 845

QY 255 EQPSRKDHKPIPSNDKELPKADTNQADKGLSPHAFKTESGETDLISPPQEEVKS 314
 Db 846 ESDS 896

QY 315 SEPTEDVEPKAEADDDTGPEEGSPKKEEKMGSGSAS-----SENREGLSDS-----TGSE 366
 Db 897 SDDS 956

QY 367 KDDLYPNGSGNGSAESS 383
 Db 957 NNVPNPSPKNGTASN 973

RESULT 3

US-09-864-761-36182
 ; Sequence 36182, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
 ; FILE REFERENCE: Aescmca-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; PRIOR FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30

Query Match 8.8%; Score 216; DB 10; Length 2344;
Best Local Similarity 25.2%; Pred. No. 0.00021;
Matches 90; Conservative 59; Mismatches 162; Indels 46; Gaps 8;

QY 22 ATESVKQEGVRSAG---NVTHPSLSORPGGS-----TKSHPEQTPKDSKSSAE 73
DB 948 STKLESQSTISLSSTDSKSMSTSESLDSTSTSDSVGSLVAGSQSVSTSDSMST 1007
QY 74 AQPDPKNSGAEAKTOKDSSKNSGAEAKTOKGTSKSGSEAEQTKDTSK--SHSELQ 131
DB 1008 SEMISDMSYSGSLAASDKSMYSVSSMSSTSGSTSESLDSTSTSDSKSLSLSTSQ 1067
QY 132 TPDKSTGKSGAEQTPDPSNRSRSGAEAKTOKDPSKSGSEAEQTKDVPKNSGADGOTPKD 191
DB 1068 SGTSTSTSTSSVRKSESOSTSGSMSTSDSTSTSTSFSDSTSDSKASTASSESISQ 1127
QY 192 GSKSGAEADTPKDVPNKSGAEAKTOKDPSKNSGAEQTPDPSKSGAEQTKDPSKPNK 251
DB 1128 SVSTSDSTSTSGSVSTSTSLSTSNERTSTSMDSSTSLSTSESDSTSDS--- 1172
QY 252 VVPEQSRKDKSPISNPSONKELPRADTNQADKGLSPHAKTESGEETDLISPPQEE 311
DB 1173 -----TSTDSISIEAISGSESTISISNESSTSDSKSASAFLESLSSES----- 1218
QY 312 VKSSEPTEDVEPREAED---DTPGPEGPPKKEEKEKMSGAS--SENREGTLSDST 363
DB 1219 --TSESTSESLSGSTSDSTSLSDSNESGCTSTSLSNSYSGSASISTSGSASTST 1273

RESULT 6
US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

Query Match 8.7%; Score 214.5; DB 10; Length 1367;
Best Local Similarity 23.1%; Pred. No. 0.00015;
Matches 96; Conservative 52; Mismatches 209; Indels 59; Gaps 16;

QY 13 AAGAVPLATESVKQEGVRSAGNVSTHPSLSORPGGS-----TKSHPEQTPK 64
DB 390 SSSAPVTPSSSTSSSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTS 449
QY 65 DSPSKSAAEQTPEDTPNKSAGAEAKTOKDSSKNSGAEAKTQKSTSKSGSEAEQTKDSTS 124

DB 450 TTES--SSAPVTPSSSTTES--SSAPVTSSTTESSSAPVTPSSSTTES--SSAPVTSSTTE 506
QY 125 KSHSELQTPKDSGK--SGAEQTPEDSPNRSRSGAEAKTQK-----DSPSKSGSEAEQ 174
DB 507 SSSAPVTPSSSTTESSSAPVTPSSSTTESSSAPVTSSTTESSSAPVTPSSSTTESSS 566
QY 175 ---TKDVPNKSGADGOTPKDSSKSGAEADTQPKDVPNKSGAEQKQ----TPKDGSKNSGA 226
DB 567 TPTVTSSTTESSSAPVTPSSSTTESSS---APVTPSSSTTESSSAPVTPSSSTTESSS 623
QY 227 -----EQGPIDGPKSGAEQTKDPSKNSKVVPEQPSRDKHGPISNPSON--KELP 276
DB 624 APVTSSTTESSSAPVTPS--SSTTESSAPV--TPSSSTTESSSAPVTPSSSTTES 679
QY 277 KADTNQADKGLSPHAKTESGEETDLISPPQEEVKSE---PTEDVEPEKAEADDDTGP 333
DB 680 SAPVTSSTTESSSAPVTSSTTESSSAPVTPSSSTTESSSAPVTP---PSSSTTESSSA 735
QY 334 EEGSPPKKEEKEKMSGASSENREG-----TLSDSTGSEKDDLYPNGSGNGSAESS 383
DB 736 PVTPSSSTTESSSAPVTSSTTESSSAPVTPSSSTTESSSAPVTPSSSTTESSS 791

RESULT 7
US-09-815-242-5898
; Sequence 5898, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5898
; LENGTH: 1349
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5898

Query Match 8.6%; Score 212; DB 10; Length 1349;
Best Local Similarity 25.4%; Pred. No. 0.0002;
Matches 54; Conservative 50; Mismatches 93; Indels 16; Gaps 3;

QY 65 DSPSKSAAEQTPEDTPNKSAGAEAKTQKDSKNSKSGAEAKTQKSTSKSGSEAEQTKDSTS 124
DB 1129 DS 1188
QY 125 KSHSELQTPKDSGKSGAEQTPEDSPNRSRSGAEAKTQKDSKNSKSGAEQTKDVPNKS 184

RESULT 13
US-09-815-242-12438
Sequence 12438, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931

```

RESULT 14
US-10-108-605-307
; Sequence 307, Application US/101086050
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQ
; TITLE OF INVENTION: PROTEINS ESSENT
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 06/176,
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 307
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; US-10-108-605-307

```

```

Query Match      8.2%; Score 202.5; DB 9; Length 724;
Best Local Similarity 21.7%; -Pred. No. 0.00034;
Matches 76; Conservative 68; Mismatches 124; Indels 83; Gaps 12;

QY 25 SVKQEEAGVRPSAGNVSTHPSLRPQGS-TKSHPEQPTPKD--SPSKSSAEAQTPETTP 81
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 SAESRHQSLSASAKSKRSRSAHRSSGARTKSGTPESQHRSGSLQNRKSGSQSRR 127

QY 82 NKSGAEAKTQKDSNKGAEAKTKQGS--TKSGSEAGTTKDTSTKSHSELQTPKQSTGK 139
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 SGSPQSRKSGTHRRGSAHRSSGARSARKSGSAODRSEKSRSHS-----GS 178

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:50:41 ; Search time 11.9604 Seconds
(without alignments)
1559.740 Million cell updates/sec

Title: US-09-763-902b-2

Perfect score: 732

Sequence: 1 MASVDFKTYDQACRAAEF.....PSNTVWKIASDCRFQDQWAS 140

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SID22/cgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID22/cgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID22/cgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID22/cgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID22/cgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID22/cgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID22/cgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID22/cgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID22/cgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID22/cgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID22/cgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID22/cgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID22/cgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID22/cgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID22/cgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID22/cgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID22/cgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID22/cgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID22/cgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID22/cgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID22/cgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID22/cgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID22/cgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	732	100.0	140	21 AAY82318	Human protein tran
2	732	100.0	164	22 AAU16228	Human novel secret
3	584	79.8	142	21 AAY82323	Human protein tran
4	584	79.8	142	23 ABB04459	Human NTF2 associa
5	584	79.8	197	22 AAB95379	Human protein sequ
6	579	79.1	162	22 AAU16227	Human novel secret
7	308	42.1	133	22 ABB65017	Drosophila melanog
8	119	16.3	688	22 ABB67404	Drosophila melanog
9	119	16.3	690	22 ABB66052	Drosophila melanog
10	118.5	16.2	125	21 AAG12073	Arabidopsis thalia

11	118.5	16.2	153	21 AAG05145	Arabidopsis thalia
12	117.5	16.1	123	21 AAG08682	Arabidopsis thalia
13	117.5	16.1	126	21 AAG08681	Arabidopsis thalia
14	117.5	16.1	131	21 AAG05147	Arabidopsis thalia
15	117.5	16.1	134	21 AAG05146	Arabidopsis thalia
16	115.5	15.8	143	21 AAY54152	Amino acid sequenc
17	105.5	14.4	130	22 ABB71855	Drosophila melanog
18	105.5	14.4	130	22 ABB71855	Mouse G3BP protein
19	105.5	14.4	477	21 AAB33532	Human cancer assoc
20	105	14.3	122	21 AAG29868	Arabidopsis thalia
21	105	14.3	157	21 AAG29867	Arabidopsis thalia
22	90	12.3	466	17 AAR95688	Human GAP-SH3 doma
23	90	12.3	466	17 AAR95688	Amino acid sequenc
24	88.5	12.1	130	22 ABB58455	Drosophila melanog
25	87.5	12.0	460	21 AAG46343	Arabidopsis thalia
26	87.5	12.0	472	21 AAG46342	Arabidopsis thalia
27	84.5	11.5	459	21 AAG20810	Arabidopsis thalia
28	84.5	11.5	471	21 AAG20809	Arabidopsis thalia
29	80.5	11.0	428	21 AAG21068	Arabidopsis thalia
30	80.5	11.0	456	21 AAG21067	Arabidopsis thalia
31	75.5	10.3	187	22 ABB60738	Drosophila melanog
32	74.5	10.2	90	21 AAG08683	Arabidopsis thalia
33	74.5	10.2	197	21 AAB18812	Amino acid sequenc
34	74	10.1	472	22 AAG04396	Novel human diagn
35	73.5	10.0	685	22 AAB30792	Amino acid sequenc
36	72.5	9.9	426	21 AAG13338	Arabidopsis thalia
37	72.5	9.9	426	21 AAG13337	Arabidopsis thalia
38	72.5	9.9	488	21 AAG13337	Arabidopsis thalia
39	72.5	9.9	488	21 AAG13336	Arabidopsis thalia
40	72.5	9.9	527	21 AAG13336	Arabidopsis thalia
41	72	9.8	178	22 ABB69707	Drosophila melanog
42	72	9.8	261	22 ABB61063	Novel human diagn
43	71.5	9.8	223	22 ABB02707	Arabidopsis thalia
44	71.5	9.8	376	21 AAG37977	Arabidopsis thalia
45	71.5	9.8	391	21 AAG37976	Arabidopsis thalia

ALIGNMENTS

RESULT 1

ID AAY82318 standard; Protein; 140 AA.

XX AC AAY82318;

XX 19-JUN-2000 (first entry)

DE Human protein transport molecule (PTM) SEQ ID NO:2.

XX Human; protein transport molecule; PTM; diagnosis; cytostatic;
KW antihypertic; antidiabetic; immunosuppressant; antitumor; osteoporotic;
KW antihypertic; antidiabetic; immunosuppressant; antitumor; osteoporotic;
KW dermatological; antidiabetic; antiparasitic; hepatotropic; antigout;
KW antinflammatory; antihiv; protein transport regulator; cancer;
KW immune disorder; cell proliferative disorder; secretory disorder;
KW urticaria; allergy; abnormal vesicle trafficking; asthma;
KW autoimmune haemolytic anaemia.

XX Homo sapiens.

XX WO200012703-A2.

XX 09-MAR-2000.

XX 26-AUG-1999; 99WO-US19616.

XX 27-AUG-1998; 98US-0098206.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;
PI Gorgone GA, Baughn MR, Patterson C;

```

XX  WPI: 2000-256642/22.
DR  N-PSDB; AAA08036.
XX
XX  New human protein transport-associated polypeptide and polynucleotide
PT  useful for diagnosis, prevention and treatment of cell proliferative
PT  and secretory disorders such as leukemia, cystic fibrosis
XX
XX  Claim 1; Page 61-62; 75pp; English.
XX
XX  AAA08035 to AAA08042 encode the human protein transport-associated
CC  molecules (PTAMs) given in AAU82317 to AAU82324. The PTAMs have
CC  cytosolic, antiarthritic, antiasthmatic, immunosuppressant,
CC  antiarthrosclerotic, antiallergic, antidiabetic, antitubercular,
CC  antirheumatic, osteopathic, dermatological, antiinfective, antiparasitic,
CC  hepatotropic, antitumor, antiinflammatory and antiHIV activities, and
CC  regulate protein transport. PTAM proteins and antagonists are useful for
CC  preventing or treating a disorder associated with decreased or increased
CC  expression or activity of PTAM. PTAM polynucleotides are useful for
CC  diagnosing conditions associated with PTAM, comprising detecting PTAM by
CC  forming a hybridisation complex, preferably after PCR amplifying the
CC  biological sample. Diseases prevented, treated or diagnosed include cell
CC  proliferative disorders such as cancers, immune disorders, secretory
CC  disorders and other conditions associated with abnormal vesicle
CC  trafficking, such as allergies, asthma, urticaria and autoimmune
CC  haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as
CC  a targeting or delivery mechanism for bringing pharmaceutical agents
CC  into cells or tissues expressing PTAM and for diagnosis of PTAM-related
CC  disorders. PTAM, its catalytic or immunogenic fragments are useful for
CC  drug screening using libraries of compounds. PTAM polynucleotides are
CC  useful for generating hybridisation probes useful in mapping the
CC  naturally occurring genomic sequences.
XX
XX  Sequence 140 AA;
XX
XX  Query Match 100.0%; Score 732; DB 21; Length 140;
XX  Best Local Similarity 100.0%; Pred. No. 2.1e-80;
XX  Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
QY 1 MASVDFKTYVDQACRAAEFVNYYITMDKRRLLSRLYMGATLWNGNAVSGQESLSE 60
DB 1 MASVDFKTYVDQACRAAEFVNYYITMDKRRLLSRLYMGATLWNGNAVSGQESLSE 60
XX
QY 61 FFEMLPSSFEQISVDDCQPVHDEATPSQTTVLVTCGSKFEGKNQKDFNQNFLTAAQS 120
DB 61 FFEMLPSSFEQISVDDCQPVHDEATPSQTTVLVTCGSKFEGKNQKDFNQNFLTAAQS 120
XX
QY 121 PSNTVYWKIASDCFRFQDWAAS 140
DB 121 PSNTVYWKIASDCFRFQDWAAS 140
XX
RESULT 2
ID AAU16228
ID AAU16228 standard; Protein; 164 AA.
AC AAU16228;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human novel secreted protein, Seq ID 1181.
XX
XX Human; immunosuppressive; antiarthritic; antirheumatic;
XX  cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
XX  neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX  vulnary; secreted protein; rheumatoid arthritis;
XX  hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
XX  cerebrovascular disorder; cerebral ischaemia; angiogenesis;
XX  nervous system disorder; Alzheimer's disease; infection; ocular disorder;
XX  corneal infection; wound healing; epithelial cell proliferation;
XX  skin ageing; food additive; preservative; antiproliferative.
XX
XX Homo sapiens.
XX

```


XX PN WO200012703-A2.
 XX PD 09-MAR-2000.
 XX PF 26-AUG-1999; 99WO-US19616.
 XX PR 27-AUG-1998; 98US-0098206.
 XX PA (INCY-) INCYTE PHARM INC.
 XX PI Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;
 XX PI Gorgone GA, Baughn MR, Patterson C;
 XX DR WPI; 2000-256642/22.
 XX DR N-PSDB; AAA08041.
 XX PT New human protein transport-associated polypeptide and polynucleotide
 XX PT useful for diagnosis, prevention and treatment of cell proliferative
 XX PT and secretory disorders such as leukemia, cystic fibrosis -
 XX PS Claim 1; Page 68; 75pp; English.
 XX CC AAA08035 to AAA08042 encode the human protein transport-associated
 XX CC molecules (PTAMs) given in AY82317 to AY82324. The PTAMs have
 XX CC cytosolic, antiarthritic, antiasthmatic, immunosuppressant,
 XX CC antiarteriosclerotic, anti-allergic, antidiabetic, antipsoriatic,
 XX CC antirheumatic, osteopathic, dermatological, antianaemic, antiproliferative,
 XX CC hepatotropic, antitumor, anti-inflammatory and anti-HIV activities, and
 XX CC regulate protein transport. PTAM proteins and antagonists are useful for
 XX CC preventing or treating a disorder associated with decreased or increased
 XX CC expression or activity of PTAM. PTAM polynucleotides are useful for
 XX CC diagnosing conditions associated with PTAM, comprising detecting PTAM by
 XX CC forming a hybridisation complex, preferably after PCR amplifying the
 XX CC biological sample. Diseases prevented, treated or diagnosed include cell
 XX CC proliferative disorders such as cancers, immune disorders, secretory
 XX CC disorders and other conditions associated with abnormal vesicle
 XX CC trafficking, such as allergies, asthma, urticaria and autoimmune
 XX CC haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as
 XX CC a targeting or delivery mechanism for bringing pharmaceutical agents
 XX CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related
 XX CC disorders. PTAM, its catalytic or immunogenic fragments are useful for
 XX CC drug screening using libraries of compounds. PTAM polynucleotides are
 XX CC useful for generating hybridisation probes useful in mapping the
 XX CC naturally occurring genomic sequences.
 XX SQ Sequence 142 AA;
 Query Match 79.8%; Score 584; DB 21; Length 142;
 Best Local Similarity 75.4%; Pred. No. 1.8e-62;
 Matches 104; Conservative 22; Mismatches 12; Indels 0; Gaps 0;
 QY 3 SVDFKTYVDQACRAAEFVNYYTMDKRRRLSLRYMGATLVNNGNAVSGQSLSEFF 62
 DB 4 SLDFKTYVDQACRAAEFVNYYTMDKRRRLSLRYMGATLVNNGNAVSGDLANNFF 63
 QY 63 EMLPSSEFQISVDCQPVHDEATPSQTVLVVIGSVKFEKGNKQDFNQNFILTAQASPS 122
 DB 64 DTLPSSEFQVNMDCQPVHEQATQSTVLVVTSGTVKFDGKQHFFNQNFLLTAQSTPN 123
 QY 123 NTWKIASDCFRFQDWS 140
 DB 124 NTWKIASDCFRFQDWS 141
 RESULT 4
 ABB04459
 ID ABB04459 standard; Protein: 142 AA.
 XX AC ABB04459;
 XX DT 04-MAR-2002 (first entry)
 XX PR

DE Human NTF2 associated protein 16.
 XX KW Human; NTF2 associated protein 16; cancer; HIV infection; gene therapy.
 XX OS Homo sapiens.
 XX PN CNI1313332-A.
 XX PD 19-SEP-2001.
 XX PF 15-MAR-2000; 2000CN-0114918.
 XX PR 15-MAR-2000; 2000CN-0114918.
 XX PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
 XX PI Mao Y, Xie Y;
 XX DR WPI; 2002-042199/06.
 XX DR N-PSDB; ABA05754.
 XX PT Polypeptide-human NTF2 associated protein 16 and polynucleotide
 XX PT encoding it -
 XX PS Claim 1; Page 25(Disclosure); 32pp; Chinese.
 XX CC The present invention provides the protein and coding sequences of human
 XX CC NTF2 associated protein 16. The sequences can be used in the treatment of
 XX CC cancer and HIV infection, as well as other diseases. The present sequence
 XX CC is the protein of the invention.
 XX SQ Sequence 142 AA;
 Query Match 79.8%; Score 584; DB 23; Length 142;
 Best Local Similarity 75.4%; Pred. No. 1.8e-62;
 Matches 104; Conservative 22; Mismatches 12; Indels 0; Gaps 0;
 QY 3 SVDFKTYVDQACRAAEFVNYYTMDKRRRLSLRYMGATLVNNGNAVSGQSLSEFF 62
 DB 4 SLDFKTYVDQACRAAEFVNYYTMDKRRRLSLRYMGATLVNNGNAVSGDLANNFF 63
 QY 63 EMLPSSEFQISVDCQPVHDEATPSQTVLVVIGSVKFEKGNKQDFNQNFILTAQASPS 122
 DB 64 DTLPSSEFQVNMDCQPVHEQATQSTVLVVTSGTVKFDGKQHFFNQNFLLTAQSTPN 123
 QY 123 NTWKIASDCFRFQDWS 140
 DB 124 NTWKIASDCFRFQDWS 141
 RESULT 5
 AAB95379
 ID AAB95379 standard; Protein: 197 AA.
 XX AC AAB95379;
 XX DT 26-JUN-2001 (first entry)
 XX DE Human protein sequence SEQ ID NO:17707.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-0116126.
 XX PR 29-JUL-1999; 99JP-0248036.
 XX PR 27-AUG-1999; 99JP-0300253.
 XX PR 11-JAN-2000; 2000JP-0118776.

OS Drosophila melanogaster.

XX WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL09120.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

XX Disclosure; SEQ ID NO 21843; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 133 AA;

Query Match 42.18; Score 308; DB 22; Length 133;

Best Local Similarity 41.48; Pred. No. 4.4e-29;

Matches 55; Conservative 27; Mismatches 47; Indels 4; Gaps 1;

Oy 5 DFKTVDOACRAAEFVNYVTMDKRRLLSRLLYMGATLVWNGNAVSGQESLSEFFEM 64

Db 4 DLKAKVESCATDTFTLLYVASVNRHQIGRLVDNATLSWNGNGAIGQMIESYFOE 63

Oy 65 LPSSFEQISVVDQVHDEATPSQTTLVVVICGSKVFEKGNKQDFNQNFILTAQASPSNT 124

Db 64 LPSSNHQNLTLDAQPIVDQAVSNQLAYLIMASGSKVFADQQLKFKQTFIVTAE----ND 119

Oy 125 VKKIASDCFRFOD 137

Db 120 KWKVSDCYRQDE 132

RESULT 8

ABB67404

ID ABB67404 standard; Protein; 688 AA.

XX ABB67404;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 29004.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC; Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL11507.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

XX Disclosure; SEQ ID NO 29004; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 688 AA;

Query Match 16.38; Score 119; DB 22; Length 688;

Best Local Similarity 30.38; Pred. No. 3.7e-05;

Matches 37; Conservative 19; Mismatches 54; Indels 12; Gaps 5;

Oy 19 EFVNVYVTMDKRRLLSRLLYMGATLVWNGNAVSGQESLSEFFEMLPSSP--FOISVW 75

Db 17 EFVROYITLLNKPMLHRRFYNNHSSYIHGESKLVVGGREHNRHQQLNFNDCHAKISQV 76

Oy 76 DCQPVDHDEATPSQTTLVVVICGSKVFEKGNKQDFNQNFILTAQASPSNTVKIASDCFRF 135

Db 77 DAQATLNG-----VVVQVTGELSNDGQPMRFRFTQTFVLAQ--SPKK--YVYHNDIFRY 127

Oy 136 QD 137

Db 128 QD 129

RESULT 9

ABB66052

ID ABB66052 standard; Protein; 690 AA.

XX ABB66052;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 24948.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PR 23-MAR-1999; 99US-0125788
PR 25-MAR-1999; 99US-0126264

b 115 RFE 117

09-MAR-19
23-MAR-19

b 115 RFE II/

09-MAR-19
23-MAR-19

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

```
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159639.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162143.
PR 29-OCT-1999; 99US-0162144.

Query Match 16.2%; Score 118.5; DB 21; Length 153;
Best Local Similarity 28.6%; Pred. No. 4.8e-06;
Matches 40; Conservative 24; Mismatches 61; Indels 15; Gaps 5;

QY 2 ASVDKFTYVDQACRAAEFVNYVYTTMDKRRRLSLYMGATLVNMGNAVSGQESLSEF 61
DB 14 ASLSKMSQMDPDVASKAFVEHYSTFDNRYGLAGLYOEASMLTFEGOKIQGVQSIKAK 73
QY 62 FEMLP--SSEFOISVVDQOPVHDEATPSOTTVLVVICGSKVEGKQR-DFNQNFIL--T 116
DB 74 LSLPFOCKKHISTVDCOP-----SGPASGMLVFVSGNLQLAGERHAKLFKSFOMFLMPT 128
QY 117 AQASPSNTVWKIASDCFRQ 136
DB 129 PQGS-----FYFNDIFSWR 143

RESULT 12
ID AAG08682 standard; Protein; 123 AA.
AC AAG08682;
DT 17-OCT-2000 (first entry)
XX
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6317.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis*thaliana.
OS
XX
XX EP1033405-A2.
XX
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
XX
XX 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
```


PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

```

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.1%; Score 117.5; DB 21; Length 126;
Best Local Similarity 30.1%; Pred. No. 4.8e-06;
Matches 37; Conservative 22; Mismatches 49; Indels 15; Gaps 5;

Qy 17 AEEFVNVTMDKRRLLSLRYMGATLVNNGNAVSGQESLSEFFMLP--SSEFQISV 74
Db 10 SKAFVHYIYSTDNRVLGAGLYQEASMLTPEGOKIOGVQSIKAKLSLPQCKHHIST 69
Qy 75 VDCOPVHDEATPSTTVLWICGSKVEGKNOR-DEFONETL--TAQASPSNTYVKIASD 131
Db 70 VDCOP-----SGPASGMLVFGVGNLQAGEHALKFSOMFLMPTPOGS-----FYVND 119

Qy 132 CFR 134
Db 120 IFR 122

RESULT 14
AAG05147
ID AAG05147 standard; Protein: 131 AA.
AC AAG05147;
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1443.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD 06-SEP-2000.
PF 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

```

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.1%; Score 117.5; DB 21; Length 131;

Best Local Similarity 29.6%; Pred. No. 5; le-06;

Matches 37; Conservative 23; Mismatches 50; Indels 15; Gaps 5;

QY 17 AEEFVNVYTTMDKRRLRLSLYMGATLVWNGNAVSGOESLSEFFEMLP--SSEFQISV 74

Db 7 SRAFEHYISTFTDNRVGLAGLYQEAASMTTFEGQIQGVOSIVAKLTSLPFOCKHHIST 66

QY 75 VDCQPVHDEATPSQTTVLVYICGSKVFEKNQR-DPQNFI--TAQASPSNTVWKIASD 131

Db 67 VDCQP-----SGPASGMVFEVSGNLQAGEKHALKFSOMFHLMPFGS----FVVFND 116

QY 132 CRRFQ 136

Db 117 IFSWR 121

RESULT 15

AAG05146

ID AAG05146 standard; Protein; 134 AA.

XX

AC AAG05146;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 1442.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140693.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142927.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

```
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.1%; Score 117.5; DB 21; Length 134;
Best Local Similarity 29.6%; Pred. NO. 5.3e-06;
Matches 37; Conservative 23; Mismatches 50; Indels 15; Gaps 5;

QY 17 AEEFVNYITMDKRRLLSRLLYMGATLWNGNAVSGQESLSEFFEMLP--SSEFQISV 74
DB 10 SKAFVEHYSTFTNVRGLAGLYQEAASMLTFEGKIOGVOSIVAKLTSLPFQOCKHHIST 69
QY 75 VDCQPVHDEATPSQTVLVVICGSVAFEGNKQR-DFNQNEIL--TAQASPSNTVWKIASD 131
DB 70 VDCQP-----SGPASCMLVFVSGNLQAGEKHALKFSQMFHLMPTQGS-----FYVFND 119
QY 132 CRRFQ 136
DB 120 IFSWR 124
```

Search completed: March 4, 2003, 14:58:32
Job time : 13.1032 secs

GenCore version 5.1.3.
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:52 ; Search time 6.29798 Seconds
(without alignments)
2137.006 Million cell updates/sec

Title: US-09-763-902B-2.

Perfect score: 732

Sequence: 1 MASVDKTYVDQACRAAEF.....PSNTWKIASDCFRFDWAS 140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	124	16.9	127	2	S00751	placental protein
2	117.5	16.1	126	2	B86405	probable nuclear t
3	116.5	15.9	125	2	S50467	nuclear transport
4	116.5	15.9	522	2	H86248	protein T23J18.22
5	107	14.6	123	2	T38039	probable nuclear t
6	106	14.5	537	2	T23479	hypothetical prote
7	105	14.3	122	2	H86398	protein F17L21.10
8	100.5	13.7	133	2	T23921	hypothetical prote
9	91.5	12.5	434	2	T40805	probable RNA-bind
10	83	11.3	427	2	E96716	probable RNA-bind
11	80.5	11.0	428	2	F86370	hypothetical prote
12	79	10.8	515	2	S63382	hypothetical prote
13	78.5	10.7	649	1	H64476	endopeptidase la h
14	77.5	10.6	398	2	I49433	gene 2B4 protein -
15	75.5	10.3	187	2	S42032	embryonal lethal (
16	74.5	10.2	197	2	A53126	syndecan-4 precurs
17	74	10.1	1044	2	T31072	heterocyst differe
18	74	10.1	1044	2	AB2158	heterocyst differe
19	73.5	10.0	108	1	K3FG	Ig light chain C r
20	73.5	10.0	685	1	EFBYS2	suppressor 2 prote
21	73.5	10.0	1940	2	T29320	myosin heavy chain
22	73	10.0	312	2	T32446	hypothetical prote
23	73	10.0	312	2	A89460	protein H42K12.1 [
24	72.5	9.9	452	2	T37899	hypothetical prote
25	72.5	9.9	488	2	D86250	hypothetical prote
26	72	9.8	579	1	WZB334	gene 34 protein -
27	71.5	9.8	453	2	T06475	probable obtusifol
28	70	9.6	509	1	S55632	deoxyribonuclease
29	70	9.6	721	2	C86667	penicillin-binding

30 70 9.6 1959 2 AG1085 hypothetical prote
31 69.5 9.5 205 2 AE1332 hypothetical prote
32 69.5 9.5 397 2 T22932 hypothetical prote
33 69 9.4 293 2 T47163 hypothetical prote
34 69 9.4 642 2 E69144 probable formate C
35 69 9.4 1227 2 T48028 hypothetical prote
36 68.5 9.4 334 2 T47693 hypothetical prote
37 68.5 9.4 667 2 B97012 methyl-accepting c
38 68.5 9.4 758 2 A98290 exopolysaccharide
39 68.5 9.4 758 2 AB2994 hypothetical prote
40 68.5 9.4 2628 2 T28651 hemagglutinin A -
41 67.5 9.2 403 2 H64861 hypothetical prote
42 67.5 9.2 2477 2 S14428 fibronectin precur
43 67 9.2 352 2 H64172 hypothetical prote
44 67 9.2 372 2 AH2214 hypothetical prote
45 67 9.2 388 2 T31705 hypothetical prote

ALIGNMENTS

RESULT 1

S00751
placental protein 15 - human
N:Alternate names: PPI5
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 24-Sep-1999
C:Accession: S00751
R:Grundmann, U.; Nerlich, C.; Rein, T.; Lottspeich, F.; Kuepper, H.A.
Nucleic Acids Res. 16, 4721, 1988
A:Title: Isolation of cDNA coding for the placental protein 15 (PPI5).
A:Reference number: S00751; MUID:88247772; PMID:3380696
A:Accession: S00751
A:Molecule type: mRNA
A:Residues: 1-127 <GRU>
A:Cross-references: EMBL:X07315; NID:g35578; PIDN:CAA30278.1; PID:g35579
C:Superfamily: yeast nuclear transport factor NTF2

Query Match 16.9%; Score 124; DB 2; Length 127;
Best Local Similarity 26.5%; Pred. No. 1.9e-05;
Matches 31; Conservative 20; Mismatches 54; Indels 12; Gaps 3;

QY 20 FVNVYTTMDKRRLLSRLLVNGTATLVWNGNAVSGOESISEFFEMLPSSFEQISVW--DC 77
DB 14 FIGHYQLFNDRTQGAIVIDASCLTWEGQFGKRAAIVKELSSLPFOKIQHSITAQDH 73
QY 78 QPVHDEATPSQTTVLVVICGVSFEGNKORDFNQNFILTAQSPSNVTWVKIASDCFR 134
DB 74 QPTPD-----SCIISMVVGQLKADEDPINGFHQMELL-----KNINDAWYCTNDMER 120

RESULT 2

B86405
probable nuclear transport factor 2 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B86405
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: B86405; MUID:21016719; PMID:11130712
A:Accession: B86405
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-126 <STO>
A:Cross-references: GB:AE005172; NID:gl1024877; PIDN:AAG36961.1; GSPDB:GN00141

C:Genetics:
A:Map position: 1
C:Superfamily: yeast nuclear transport factor NTF2

Query Match 16.1%; Score 117.5; DB 2; Length 126;
Best Local Similarity 30.1%; Pred. NO. 8.5e-05;
Matches 37; Conservative 22; Mismatches 49; Indels 15; Gaps 5;

QY 17 AEEFYNVVYTTMDKRRLLSLRYMGTATLVWNGNAVSGESLSEFFEMLP--SSFEQISV 74
DB 10 SKAFVHYHSYFTDNRVGLAGLQASMLTFEGQIQGVIVAKLTSLPFOQCKHHIST 69
QY 75 VDCQPVHDEATPSQITVLVWICGSKVKEGKQK--DFNQNFI--TAQASPSNTVWIKIASD 131
DB 70 VDCQP-----SGPASGLVFSGNLQAGEEHALKFSQNFHMLPTPQGS-----FYFEND 119
QY 132 CFR 134
DB 120 IFR 122

RESULT 3
S50467
nuclear transport factor NTF2 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YER009w
C:Species: Saccharomyces cerevisiae
C:Date: 28-Jan-1995 #sequence.Revision 12-May-1995 #text_change 24-Sep-1999
C:Accession: S50467; S41793; S72237
R:Dietrich, F.S.
Submitted to the EMBL Data Library, December 1994
A:Description: The sequence of *S. cerevisiae* cosmids 9537, 9581, 9495, 9867, and lambda
A:Reference number: S50459
A:Accession: S50467
A:Molecule type: DNA
A:Molecule type: DNA
A:Residues: 1-125 <DIE>
A:CROSS-references: EMBL:U18778; NID:g603592; PIDN:AAB64542.1; PID:g603601; MIPS:YER009w
R:Haarer, B.K.; Petzold, A.S.; Brown, S.S.
submitted to the EMBL Data Library, July 1993
A:Description: Identification of mutations that are synthetically lethal with altered ye
A:Reference number: S41793
A:Accession: S41793
A:Molecule type: DNA
A:Residues: 1-84 <HAA>
A:CROSS-references: EMBL:L22204; NID:g347714; PIDN:AAB49379.1; PID:g347715
R:Haarer, B.K.; Corbett, A.; Kweon, Y.; Petzold, A.S.; Silver, P.; Brown, S.S.
submitted to the EMBL Data Library, July 1993
A:Title: SEC3 mutations are synthetically lethal with profilin mutations and cause defec
A:Reference number: S72237; NUID:9704444; PMID:8889515
A:Accession: S72237
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-84 <HAW>
A:CROSS-references: EMBL:L22204; NID:g347714; PIDN:AAB49379.1; PID:g347715
C:Genetics:
A:Gene: SGD:NTF2
A:CROSS-references: SGD:S0000811; MIPS:YER009w
A:Map position: 5R
C:Superfamily: yeast nuclear transport factor NTF2

Query Match 15.9%; Score 116.5; DB 2; Length 125;
Best Local Similarity 30.9%; Pred. NO. 0.00011;
Matches 42; Conservative 19; Mismatches 54; Indels 21; Gaps 7;

QY 3 SVDFKYYVDQACRAAEFYNVYTTMDKRRLLSLRYMGTATLVWNGNAVSGESLSEFF 62
DB 2 SLDFNT-----LAQNTQFYNQFDTDRSQLGNLYRNESMLTFTESLQAGAKDIVEKL 54
QY 63 EMLPSSEFO--ISVDCQPVHDEATPSQITVLVWICGS--VKFEGNKQKRDFFNQNFILTAQ 118
DB 55 VSLPFOKXVQRITLDAQP-----ASPN-GDVLVYMITGDLILDEQNQR-FSQVPHLI-- 106
QY 119 ASPSNTVWIKIASDCFR 134

A;Cross-references: GB:AE005172; NID:g9802547; PIDN:AAF99749.1; GSPDB:GN00141
C:Genetics:
A:Gene: F17L21.10
A:Map position: 1
C:Superfamily: yeast nuclear transport factor NTF2

Query Match 14.3%; Score 105; DB 2; Length 122;
Best Local Similarity 27.3%; Pred. No. 0.0016;
Matches 33; Conservative 22; Mismatches 54; Indels 12; Gaps 4;

QY 17 AEEFVNVYYTMDKRRLLSRLYMGATLVVNGNAVSGESLSEFFEMLP--SSEFOISV 74
DB 7 AKAFVEHYSTFDANRPGVLSLYQESMLFEGKIOGSONIVAKLTGLFPQCKHNIT 66

QY 75 VDCOPVDEATPSTQTVLVVVCVSGVFEGNKQR-DFNQNFILTAQAQSPSNVWKAISDCF 133
DB 67 VDCOP-----SGPAGGMLVFVSGNLQAGEQHAKFKSQMFHLISNQGN---YVVFNDIF 117

QY 134 R 134
DB 118 R 118

RESULT 8
T23921
hypothetical protein R05D11.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T23921
R;White, S.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19818
A:Accession: T23921
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-133 <WIL>
A:Cross-references: EMBL:Z75546; PIDN:CAA99890.1; GSPDB:GN00019; CESP:R05D11.3
A:Experimental source: clone R05D11
C:Genetics:
A:Gene: CESP:R05D11.3
A:Map position: 1
A:Introns: 33/1; 63/3; 96/3
C:Superfamily: yeast nuclear transport factor NTF2

Query Match 13.7%; Score 100.5; DB 2; Length 133;
Best Local Similarity 26.8%; Pred. No. 0.005;
Matches 34; Conservative 29; Mismatches 45; Indels 19; Gaps 6;

QY 17 AEEFVNVYYTMD-----KRRLLSRLY-MGATPLVWNGNAVSGESLSEFFEMLPSEF 70
DB 11 AKAFIQHYYSKFDVGDGMSRAQGLSDLYDPENSYMTFEGQAQKRGDILQKFTTLGFTKI 70

QY 71 Q--ISVVDQCPVHDEATPSTQTVLVVVCVSGVFKFGNKQRDNQNFILTAQAQSPNT-VWK 127
DB 71 QRAITVTDSPLOYD-----GSIQVWGLQKLTDEDPINPESQVFIL---RPNNQGSYF 120

QY 128 IASDCFR 134
DB 121 IGNEIFR 127

RESULT 9
T40805
probable RNA-binding protein - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40805
R;Beck, A.; Reinhardt, R.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z21949
A:Accession: T40805
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-434 <BEC>
A:Cross-references: EMBL:AL032684; PIDN:CAA21796.1; GSPDB:GN00067; SPDB:SPBP8B7.11
A:Experimental source: strain 972h; clone pl p8B7
C:Genetics:
A:Gene: SPDB:SPBP8B7.11
A:Map position: 2
A:Introns: 38/3; 63/3

Query Match 12.5%; Score 91.5; DB 2; Length 434;
Best Local Similarity 27.1%; Pred. No. 0.17;
Matches 32; Conservative 18; Mismatches 61; Indels 7; Gaps 4;

QY 20 FVNVYTTMDKRRLLSRLYMGATLW--NGNAVSGQSLSEFFEMLPSSSEFOISVDC 77
DB 22 FVEYTYLKNPRLNHCYTKSLIHGDEGESLCHGQOEIHNKILDLDFONCKVLI 81

QY 78 QPVHDEATPSQTTVLVVICGVSVKFEGNKQDNQNFILTAQASPSNTVWVKIADCFRF 135
DB 82 SNV-DSLASSNGGIVQVLGEMSNKGLSRKFAQTFFLAEQ--PNG--YFVLNDIFRF 134

RESULT 10
E96716
probable RNA-binding protein F23010.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E96716
R:Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marziall,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E96716
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-427 <STO>
A:Cross-references: GB:AE005173; NID:g7705097; PIDN:AAF67776.1; GSPDB:GN00141
C:Genetics:
A:Gene: F23010.17
A:Map position: 1

Query Match 11.3%; Score 83; DB 2; Length 427;
Best Local Similarity 25.2%; Pred. No. 1.2;
Matches 35; Conservative 27; Mismatches 51; Indels 26; Gaps 6;

QY 10 VQACRAAEFVNVYTTMDKRRLLSRLYM-----GTATLVWNGNAVSGQESL 58
DB 7 VPSAODIAAEFVQYHYHVLGQLPHEARLYVDASVSRPDVTGTWMSFTSVEAIN-KHIL 65

QY 59 SEFFEMLPSSSEFOISVDCQPVHDEATPSQTTVLVVICGVSVKFEGNKQDNQNFILTAQ 118
DB 66 SCDFE---NFKFVLVSDVNSQSLDEG-----IFIMVIGMTGKNRRKFSOMFYLAQ 116

QY 119 ASPSNTVWVKIADCFRFQD 137
DB 117 ---NTL-VVLNDMLRYVD 130

RESULT 11
F86270
hypotheical protein F2123.16 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: F86270
R:Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marziall,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E96716
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-427 <STO>
A:Cross-references: GB:AE005173; NID:g7705097; PIDN:AAF67776.1; GSPDB:GN00141
C:Genetics:
A:Gene: F23010.17
A:Map position: 1

Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marziall,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: F86270
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <STO>
A:Cross-references: GB:AE005172; NID:g8920577; PIDN:AAF81299.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 11.0%; Score 80.5; DB 2; Length 428;
Best Local Similarity 25.8%; Pred. No. 2.2;
Matches 31; Conservative 26; Mismatches 52; Indels 11; Gaps 6;

QY 20 FVNVYTTMDKRRLLSRLYMGATLWNGN--AVSGQSLSEFFEMLPSSSEFOISVDC 77
DB 19 FVEKYNLYKSPSQVHGFYLDSDVLRPGSGGEMVSVKSLKAINEQIMSFDEISKIOI 78

QY 78 QPVHDEATPSQTTVLVVICG--SVKFEKNQKQDNQNFILTAQASPSNTVWVKIADCFRF 135
DB 79 LTADSQAS-YMNGVTVLTGLTVK-EGQMR-FSQSFLV---PLNGSYFVLNDVFRY 131

RESULT 12
S63382
hypotheical protein YNR051c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypotheical protein N3465
C:Species: Saccharomyces cerevisiae
C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
C:Accession: S63382; S27439; S27440
R:Pohl, T.M.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63346
A:Accession: S63382
A:Molecule type: DNA
A:Residues: 1-515 <POH>
A:Cross-references: EMBL:Z71666; NID:gl302565; PID:g1302566; MIPS:YNR051
A:Experimental source: strain S288C
R:Cusick, M.E.
submitted to the EMBL Data Library, March 1992
A:Reference number: S27437
A:Accession: S27439
A:Molecule type: DNA
A:Residues: 1-316 'T' <CUS>
A:Cross-references: EMBL:M88607; NID:gl72079; PID:gl72081
A:Accession: S27440
A:Molecule type: DNA
A:Residues: 350-441 <CUW>
A:Cross-references: EMBL:M88607; NID:gl72079; PID:g553136
C:Genetics:
A:Gene: SGD:BRES
A:Cross-references: SGD:S0005334
A:Map position: 14R

Query Match 10.8%; Score 79; DB 2; Length 515;
Best Local Similarity 23.0%; Pred. No. 3.9;
Matches 34; Conservative 24; Mismatches 56; Indels 34; Gaps 6;

QY 10 VQACRAAEFVNVYTTMDKRRLLSRLYMGATLWNGN-----AVSG 54
DB 5 VQDICFA---FLQNYERMRTDPSKLAYFYASTAELTHNYQSKSTNEKDDVLPVTKVFG 61

QY 55 QESLSSEFFE-----MLPSSSEFOISVDCQ---PVHDEATPSQTTVLVVICGVSFKFENKOR 107
DB 62 RENINKFFSRNDKAKVSLKLDITDFOYTGHLH-----KSLIMATGEMFTGTTPVY 114

QY 108 DFNQNFILTAQASPSNTVWVKIADCFRF 135

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:51:02 ; Search time 3.35122 Seconds
(without alignments)
1732.709 Million cell updates/sec

Title: US-09-763-902B-2
Perfect score: 732
Sequence: 1 MASVDFKTYVDQACRAAEF.....PSNTVVKIASDCFRFQDWS 140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %	Score	Match	Length	DB ID	Description
1	732	100.0	140	1	NXT1_HUMAN	Q9UKK6 homo sapien
2	726	99.2	140	1	NXT1_MOUSE	Q9NPJ8 homo sapien
3	584	79.8	142	1	NXT2_HUMAN	Q9V3H8 drosophila
4	308	42.1	133	1	NXT1_DROME	Q9U757 caenorhabdi
5	206.5	28.2	137	1	NXT1_CAEEL	P13662 homo sapien
6	124	16.9	127	1	NXT2_HUMAN	O42242 xenopus lae
7	123	16.8	127	1	NXT2_YENLA	Q9P926 candida alb
8	118.5	16.2	124	1	NXT2_CANAL	Q9C7F5 arabidopsis
9	117.5	16.1	126	1	NXT2_ARATH	P33331 saccharomyc
10	116.5	15.9	125	1	NXT2_YEAST	Q9UN86 homo sapien
11	115.5	15.8	482	1	G3B2_HUMAN	P97379 mus musculu
12	115.5	15.8	482	1	G3B2_MOUSE	Q9X154 oryza sativ
13	107	14.6	122	1	NXT2_ORYSA	Q10100 schizosacch
14	107	14.6	123	1	NXT2_SCHPO	Q13283 homo sapien
15	105.5	14.4	466	1	G3BP_HUMAN	P87102 neurospora
16	102.5	14.0	124	1	NXT2_NEUCR	P97855 mus musculu
17	101.5	13.9	465	1	G3BP_MOUSE	O21735 caenorhabdi
18	100.5	13.7	133	1	NXT2_CAEEL	Q94260 schizosacch
19	91.5	12.5	434	1	G3BP_SCHPO	P53741 saccharomyc
20	79	10.8	515	1	YN87_YEAST	Q58812 methanococc
21	78.5	10.7	649	1	LNH1_METJA	P82147 drosophila
22	75.5	10.3	187	1	EFL2_DROME	P49416 gallus gall
23	74.5	10.2	197	1	SDC4_CHICK	P11272 rana catesb
24	73.5	10.0	108	1	KAC_RANCA	P05453 saccharomyc
25	73.5	10.0	685	1	ERF2_YEAST	P02565 gallus gall
26	73.5	10.0	1940	1	MYH3_CHICK	P09287 varicella-z
27	72	9.8	579	1	UL25_VZVD	P93596 triticum ae
28	71.5	9.8	453	1	CP51_WHEAT	P09004 zea mays (m
29	69.5	9.5	219	1	MI25_MAIZE	P58210 rhizobium l
30	69	9.4	273	1	DABL_RHIZO	Q00762 oxytricha g
31	68.5	9.4	559	1	TCPG_OXYGR	Q51845 porphyromon
32	68.5	9.4	2628	1	HAGA_PORGI	P75990 escherichia
33	67.5	9.2	403	1	YCGF_ECOLI	

34	67.5	9.2	2477	1	FINC_MOUSE	P11276 mus musculu
35	67.5	9.2	2477	1	FINC_RAT	P04937 rattus norv
36	67	9.2	352	1	YGIF_HAEIN	P45267 haemophilus
37	67	9.2	3767	1	MUA3_CAEEL	P34576 caenorhabdi
38	66.5	9.1	433	1	ERF1_SCHPO	P79063 schizosacch
39	66.5	9.1	563	1	YWO3_CAEEL	Q10906 caenorhabdi
40	66.5	9.1	569	1	ILIR_HUMAN	P14778 homo sapien
41	66	9.0	268	1	DAPB_BRUME	Q8YD68 bruceella me
42	66	9.0	559	1	TCPG_TETPY	P54408 tetrahymena
43	65.5	8.9	492	1	CP51_SORBI	P93846 sorghum bic
44	65.5	8.9	1940	1	MYH3_HUMAN	P11055 homo sapien
45	65.5	8.9	1940	1	MYH3_RAT	P12847 rattus norv

ALIGNMENTS

RESULT 1
ID NXT1_HUMAN STANDARD; PRT; 140 AA.
AC Q9UKK6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE NTF2-related export protein 1 (p15 protein).
GN NTF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=20036817; Pubmed=10567585;
RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;
RT "Identification of an NTF2-related factor that binds Ran-GTP and
RT regulates nuclear protein export.";
RL Mol. Cell. Biol. 19:8616-8624(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; Pubmed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavridis G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Connor R.E.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Gwilliam R.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammon S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharaslo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McKay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung, Placenta, and Uterus;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [4]

RP FUNCTION:
RX MEDLINE-20307861; PubMed-10848583;
RA Ossareh-Nazari B., Maison C., Black B.E., Levesque L., Paschal B.M.,
RA Dargemont C.;
RT "RanGTP-binding protein NX1L facilitates nuclear export of different
RT classes of RNA in vitro";
RL Mol. Cell. Biol. 20:4562-4571(2000).
RN [5]
RN FUNCTION:
RX MEDLINE-21160285; PubMed-11259602;
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,
RA Paschal B.M., Rekosh D., Hammarshjold M.L.;
RT "NX1L (p15) is a crucial cellular cofactor in TAP-dependent export of
RT intron-containing RNA in mammalian cells";
RL Mol. Cell. Biol. 21:2545-2554(2001).
RN [5]
CC -!- FUNCTION: Stimulator of protein export for NES-containing
CC proteins. Also plays a role in the nuclear export of U1 snRNA,
CC tRNA, and mRNA.
CC -!- SUBUNIT: Preferentially binds Ran-GTP. Associates with NX1L, NXF2,
CC NXF3 and NXF5.
CC -!- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and
CC the cytoplasm.
CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; AF156957; AAD54942.1;
CC EMBL; AL096677; CAC03437.2;
CC EMBL; BC000759; AAH00759.1;
CC EMBL; BC002687; AAH02687.1;
CC EMBL; BC003029; AAH03029.1;
CC EMBL; BC003410; AAH03410.1;
CC Genew; HGNC:15913; NNT1.
CC MIM; 605811;
CC InterPro; IPR002075; NTF2.
CC Pfam; PF02136; NTF2; 1.
CC PROSITE; PS50177; NTF2_DOMAIN; 1.
CC Transport; Protein transport; Nuclear protein.
FT DOMAIN 16 135 NTF2
SQ SEQUENCE 140 AA; 15847 MW; 358FA86AC3944594 CRC64;
Query Match 100.0%; Score 732; DB 1; Length 140;
Best Local Similarity 100.0%; Pred. No. 5.3e-69;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASVDFKTYVDQACRAAEFVNVYTTMDKRRLLSRLYMGATLVWNGNAVSGQESLSE 60
DB 1 MASVDFKTYVDQACRAAEFVNVYTTMDKRRLLSRLYMGATLVWNGNAVSGQESLSE 60
QY 61 FFEMLPSEFOISVVDQCPVHDEATPSQTTVLVVICGSKVFEKGNKQDFNQNFLTAAQS 120
DB 61 FFEMLPSEFOISVVDQCPVHDEATPSQTTVLVVICGSKVFEKGNKQDFNQNFLTAAQS 120
QY 121 PSNTVWKIASDCRFQDQWAS 140
DB 121 PSNTVWKIASDCRFQDQWAS 140
RESULT 2
ID NX1L_MOUSE STANDARD; PRT; 140 AA.
AC Q9QZV9; Q9DIP7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE NTF2-related export protein 1.
GN NX1L.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID-10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20036817; PubMed-10567585;
RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;
RT "Identification of an NTF2-related factor that binds Ran-GTP and
RT regulates nuclear protein export";
RL Mol. Cell. Biol. 19:8616-8624(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-21085660; PubMed-11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Willing L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: Stimulator of protein export for NES-containing
CC proteins. Also plays a role in the nuclear export of U1 snRNA,
CC tRNA, and mRNA (by similarity).
CC -!- SUBUNIT: Preferentially binds Ran-GTP. Associates with NX1L, NXF2
CC and NXF3.
CC -!- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and
CC the cytoplasm.
CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; AF156958; AAD54943.2;
CC EMBL; AK003254; BAB22670.1;
CC MGD; MGI:1929619; Nxtl.
CC InterPro; IPR002075; NTF2.
CC Pfam; PF02136; NTF2; 1.
CC PROSITE; PS50177; NTF2_DOMAIN; 1.
CC Transport; Protein transport; mRNA transport; Nuclear protein.
FT DOMAIN 16 135 NTF2.
FT CONFLICT 100 100 K -> N (IN REF. 2).
SQ SEQUENCE 140 AA; 15847 MW; 358EAFDC19FE4594 CRC64;
Query Match 99.2%; Score 726; DB 1; Length 140;
Best Local Similarity 98.6%; Pred. No. 2.2e-68;
Matches 138; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASVDFKTYVDQACRAAEFVNVYTTMDKRRLLSRLYMGATLVWNGNAVSGQESLSE 60
DB 1 MASVDFKTYVDQACRAAEFVNVYTTMDKRRLLSRLYMGATLVWNGNAVSGQESLSE 60
QY 61 FFEMLPSEFOISVVDQCPVHDEATPSQTTVLVVICGSKVFEKGNKQDFNQNFLTAAQS 120
DB 61 FFEMLPSEFOISVVDQCPVHDEATPSQTTVLVVICGSKVFEKGNKQDFNQNFLTAAQS 120

```

DB 61 PFEMPSSEFOISVDDCPVHDDATPSQTTVLVWICGTVEGKQDNQNFILTAQAS 120
QY 121 PSNTVWKIASDCFRFDQWAS 140
DB 121 PSNTVWKIASDCFRFDQWAS 140

RESULT 3
NXT2_HUMAN
ID NXT2_HUMAN STANDARD; PRT; 142 AA.
AC Q9NPJ8; Q9H8U0; Q9NRL7; Q9Y3M4; Q9Y3M5; Q9NQ64;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE NTF2-related export protein 2 (p15-2 protein) (DC9) (BM025).
GN NXT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Testis;
RX MEDLINE=20528640; PubMed=11073998;
RA Herold A., Suyama M., Rodrigues J.P., Braun I.C., Kutay U.,
RA Carmo-Fonseca C., Bork P., Izaurralde E.;
RT "TAP (NXF1) belongs to a multigene family of putative RNA export
RT factors with a conserved modular architecture.";
RL Mol. Cell. Biol. 23:8996-9008(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Kang Y., Cullen B.;
RT "p15-2, a homologous protein of p15, interacts with Tap.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Li Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.;
RT "Novel genes expressed in human dendritic cell.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Ovarian carcinoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Negai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Bone marrow;
RA Zhao M., Song H., Li N., Peng Y., Han Z., Chen Z.;
RT "A novel gene expressed in human bone marrow.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA Howden P.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Regulator of protein export for NES-containing proteins.
CC Also plays a role in mRNA nuclear export.
CC -1- SUBUNIT: Associates with NXF1, NXF2, NXF3 and NXF5.
CC -1- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and
CC the cytoplasm.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; A (shown here) and B; are
CC produced by alternative splicing.
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ277591; CAB96371.1; -
DR EMBL; AJ278323; CAC01129.1; -
DR EMBL; AF246127; AAF78034.1; -
DR EMBL; AF201942; AAF68878.1; -
DR EMBL; AK032289; BAB14511.1; ALT_INIT.
DR EMBL; AF212223; AAF87325.1; ALT_INIT.
DR EMBL; AL031387; CAB41301.1; -
DR EMBL; AL031387; CAB41302.1; -
DR MIM; 300320; -
DR InterPro; IPR002075; NTF2.
DR Pfam; PF02136; NTF2; 1.
DR PROSITE; PS0177; NTF2_DOMAIN; 1.
KW Transport; Protein transport; mRNA transport; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 17 136 NTF2.
FT VARSPPLIC 1 5 MATSL -> VTNHVPALCTAGRGPRFAARPAGPTPFLVSSR
FT FLIPERTLSWEIARGGDVG (IN ISOFORM B).
SQ SEQUENCE 142 AA; 16228 MW; 009B4E8929A6BA6C CRC64;

Query Match 79.8%; Score 584; DB 1; Length 142;
Best Local Similarity 75.4%; Pred. No. 1.2e-53;
Matches 104; Conservative 22; Mismatches 12; Indels 0; Gaps 0;

QY 3 SVDFKTYVDQACRAAEFEFVNYVTMDKRRRLSLYMGATLVWNGNAVSGQESLSEFF 62
DB 4 SLDFKTYVDQACRAAEFEFVNYVTMDKRRRLSLYMGATLVWNGNAVSGQESLSEFF 63
QY 63 EMLPSEFOISVDDCPVHDDATPSQTTVLVWICGTVEGKQDNQNFILTAQASPS 122
DB 64 DTLPSSEFOVNMDCQPVHEQATQSTTVLVVTSCTVAFDGNKQHFNFQNFILTAQSTPN 123
QY 123 NTWVKIASDCFRFDQWAS 140
DB 124 NTWVKIASDCFRFDQWSS 141

RESULT 4
NXTL_DROME
ID NXTL_DROME STANDARD; PRT; 133 AA.
AC Q9V3H8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE NTF2-related export protein (p15).
GN NXT1 OR CGI2752.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20036817; PubMed=10567585;
RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;
RT "Identification of an NTF2-related factor that binds Ran-GTP and
RT regulates nuclear protein export.";
RL Mol. Cell. Biol. 19:8616-8624(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

```

```

RESULT 5
NNXTL_CAEEL STANDARD; PRT; 137 AA.
Q9U757;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
NTF2-related export protein.
NXTI OR Y7IF9AM.5.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditoidea;
Rhabditiidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;
[1]
SEQUENCE FROM N.A.
MEDLINE=20036817; PubMed=10567585;
Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;
"Identification of an NTF2-related factor that binds Ran-GTP and
regulates nuclear protein export.";
Mol. Cell. Biol. 19:8616-8624(1999).
[2]
SEQUENCE FROM N.A.
STRAIN=Bristol N2;
Bradshaw-Cordum H., Scott K., Graves T.;
Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
!- FUNCTION: Stimulator of protein export for NES-containing
proteins. Also plays a role in the nuclear export of, U1 snRNA,
tRNA, and mRNA (by similarity).
!- SUBUNIT: Preferentially binds Ran-GTP (By similarity).
!- SUBCELLULAR LOCATION: Nuclear (By similarity).
!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF156960; RAD54945.1; -
EMBL; AC024852; AAK66028.1; -
HSP; P13662; 1OUN.
Wormpep; X71F9AM.5; CE26780.
InterPro; IPR002075; NTF2.
Pfam; PF02136; NTF2; 1.
PROSITE; PS50177; NTF2_DOMAIN; 1.
Transprot; Protein transport; mRNA transport; Nuclear protein.
DOMAIN 19 135 NTF2.
SEQUENCE 137 AA; 15479 MW; B99ABE3CA3479962 CRC64;

Query Match      28.2%; Score 206.5; DB 1; Length 137;
Best Local Similarity 31.2%; Pred. NO. 1.6e-14;
Matches 40; Conservative 35; Mismatches 48; Indels 5; Gaps 2;

QY 11 DOACRAAEFEVNVNYTMDKRRLLSRLYNGTATLVNNGNAVSGOESLSSEFFEMLPSEF 70
   : : : :::::||||| ||||: : : : : ||||: : : | : : |||: :
Db 14 EELCNKKFMVDYYDMDKRKREKIGFLYTQVNVAWNGNPINGYSICEFMKALPSTQH 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 71 QISVDCQPQVDEATPSQT-TVLVLTGCGVKFGKNKORDENQNFIATAQASPNVTKIA 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 DIQLSDAQRLEPGVTGDMGGMLLNAGAVTVGDGSKRAFTQILLGVE----DGKFKVK 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 130 SDCFRFD 137
   || ||: |
Db 130 SDRFERYD 137

RESULT 6
NTF2_HUMAN STANDARD; PRT; 127 AA.
AC P13662;
```

DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Nuclear transport factor 2 (NTF2) (Placental protein 15) (PPI15).
 GN NTF2 OR PPI15.
 OS Homo sapiens (Human),
 OS Mus musculus (Mouse), and
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606, 10090, 10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Human;
 RA MEDLINE=88247772; PubMed=3380696;
 RX Grundmann G., Nerlich C., Rein T., Lottspeich F., Kuepper H.A.;
 RT "Isolation of cDNA coding for the placental protein 15 (PPI15).";
 RL Nucleic Acids Res. 16:4721-4721(1988).
 RN [2]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RC SPECIES=Human;
 RA MEDLINE=95263674; PubMed=7744965;
 RX Paschal B.M., Gerace L.;
 RT "Identification of NTF2, a cytosolic factor for nuclear import that
 RT interacts with nuclear pore complex protein p62.";
 RL J. Cell Biol. 129:925-937(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Human, and Mouse;
 RA Strausberg R.;
 RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Mouse; STRAIN=C57BL/6J; TISSUE=Kidney;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Aono H., Arai A.,
 RA Arakawa T., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanaoka T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kato H., Kawai J., Kojima Y.,
 RA Konno H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Okazaki Y., Okido T.,
 RA Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D.,
 RA Shibata K., Shibata Y., Shingawa A., Shiraki T., Sogabe Y.,
 RA Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y.,
 RA Toyota T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Rat; TISSUE=Kidney;
 RA Kent H.M., Clarkson W.D., Bullock T.L., Stewart M.;
 RT Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RC SPECIES=Human;
 RA MEDLINE=96300274; PubMed=8757804;
 RX Bullock T.L., Clarkson W.D., Kent H.M., Stewart M.;
 RT "The 1.6-A resolution crystal structure of nuclear transport factor 2
 RT (NTF2).";
 RL J. Mol. Biol. 260:422-431(1996).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RA MEDLINE=9805028; PubMed=9368653;
 RX Clarkson W.D., Corbett A.H., Paschal B.M., Kent H.M., McCoy A.J.,
 RA Gerace L., Silver P.A., Stewart M.;
 RT "Nuclear protein import is decreased by engineered mutants of nuclear
 RT transport factor 2 (NTF2) that do not bind GDP-Ran.";
 RL J. Mol. Biol. 272:716-730(1997).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF COMPLEX WITH RAN.
 RC SPECIES=Rat;
 RA MEDLINE=98202576; PubMed=9533885;
 RX Stewart M., Kent H.M., McCoy A.J.;
 RT "Structural basis for molecular recognition between nuclear transport

RT factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase
 RT Ran.";
 RL J. Mol. Biol. 277:635-646(1998).
 CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS.
 CC INTERACTS WITH THE NUCLEOPORIN P62 AND WITH RAN. ACTS AT A
 CC RELATIVELY LATE STAGE OF NUCLEAR PROTEIN IMPORT, SUBSEQUENT TO THE
 CC INITIAL DOCKING OF NUCLEAR IMPORT LIGAND AT THE CYTOSOLIC ENVELOPE.
 CC COULD BE PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT
 CC ASSEMBLE AT THE PORE COMPLEX DURING NUCLEAR IMPORT.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X07315; CA030278.1;
 DR EMBL; X91651; CAAG2839.1;
 DR EMBL; BC003955; AAH03955.1;
 DR EMBL; BC002348; AAH02348.1;
 DR EMBL; AK002461; BAB22117.1;
 DR EMBL; AK012500; BAB28283.1;
 DR EMBL; AK020512; BAB32122.1;
 DR EMBL; UA3939; AAAS5905.1;
 DR PIR; S00751; S00751.
 DR PDB; 1OUN; 01-APR-97.
 DR PDB; 1ARO; 24-DEC-97.
 DR PDB; 1ASK; 28-JAN-98.
 DR PDB; 1AZK; 29-APR-98.
 DR MIM; 605813;
 DR MGD; MGI:895065; Ntf2.
 DR InterPro; IPR002075; NTF2.
 DR Pfam; PF02136; NTF2; 1.
 DR PROSITE; PS01177; NTF2_DOMAIN; 1.
 DR TRANSPORT; Protein transport; 3D-structure.
 KW DOMAIN 10 121 NTF2.
 FT 'DOMAIN 10 121 NTF2'.
 SQ SEQUENCE 127 AA; 14478 MW; 817752F20E262FD3 CRC64;
 Query Match 16.9%; Score 124; DB 1; Length 127;
 Best Local Similarity 26.5%; Pred. No. 5.4e-06;
 Matches 31; Conservative 20; Mismatches 54; Indels 12; Gaps 3;
 QY 20 FVNVYTTMDKRRLLSRLYMGATLVNMGNAVSGQSLSEFFEMLPSEFQISVW--DC 77
 Db 14 FIQHYQLFDNDRTQLGAIYIDASCLTWEGQFOGKAAIVEKLSLPPFKIQHSITAQDH 73
 QY 78 QPVHDEATPSTQTVLVVVICGSKFEGNKQRDNFNQNTLTAQASPSNTVWKTASDCFR 134
 Db 74 QPTPD-----SCIISWVVGOLKADDPIMGFHOMFL-----KNINDAWVCTNDMFR 120
 RESULT 7
 NTF2_XENLA STANDARD; PRT; 127 AA.
 ID NTF2_XENLA
 AC O42242;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nuclear transport factor 2 (NTF-2) (P10).
 GN NTF2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pu R.T., Shelsta H.N., Dasso M.;

TO: NATIONAL DEFENSE ACADEMY, 1000 PENTAGON DRIVE, ARLINGTON, VA 22204-4501

RL DNA Res. 5:169-176(1998).
 RP [3] SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=Brain;
 RA Guillard E.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=Brain;
 RA Kennedy D., Ru K., Mattick J.S.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=B-cell;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [6]
 RP INTERACTION WITH IKAPPABALPHA.
 RX MEDLINE-20549669; PubMed-10969074;
 RA Prigent M., Barlat I., Langan H., Dargemont C.;
 RT "Ikappabalpha and Ikappabalpha/NF-kappa B complexes are retained in the cytoplasm through interaction with a novel partner, RasGAP SH3-binding protein 2.";
 RL J. Biol. Chem. 275:36441-36449(2000).
 CC -!- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA TRANSPORT (POTENTIAL).
 CC -!- SUBUNIT: BINDS TO THE N-TERMINAL DOMAIN OF IKAPPABALPHA.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 DR EMBL; AF145284; AAC53553.1; -
 DR EMBL; AB014560; BAA31635.1; -
 DR EMBL; AF051311; AAC15705.1; -
 DR EMBL; AF053535; AAC95292.1; -
 DR EMBL; BC011731; AAH11731.1; -
 DR HSSP; P09651; 1HA1.
 DR InterPro; IPR002075; NTF2.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; PF02136; NTF2; 1.
 DR PROSITE; PS00360; RRM; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 KW Transport; RNA-binding; Alternative splicing.
 FT DOMAIN 11 133 NTF2.
 FT DOMAIN 331 409 RNA-BINDING (RRM).
 FT DOMAIN 134 223 GLU-RICH.
 FT DOMAIN 419 479 GLY-RICH.
 FT VARSPPLIC 243 275 MISSING (IN ISOFORM B).
 FT CONFLICT 267 267 S -> P (IN REF. 2).
 FT CONFLICT 359 359 E -> V (IN REF. 3).
 FT CONFLICT 460 460 M -> I (IN REF. 3).
 SQ SEQUENCE 482 AA; 54111 MW; 2545CGA3FAAE218 CRC64;
 Query Match 15.8%; Score 115.5; DB 1; Length 482;
 Best Local Similarity 29.9%; Pred. No. 0.00021;
 Matches 38; Conservative 18; Mismatches 58; Indels 13; Gaps 4;
 QY 19 EFNWYVYTTMDKRRLLRLSLRYMGTATLVNNG-----NAVSGQSLSEFFEMLPSSSE 70
 DB 14 EFRVQYITLLNKAPEYLRHYRFGNRSYVHGVDASGKPOENVYQGDIIHKVLSLNFSEC 73

QY 71 QISVVDQCPVHDEATPSQTTLVLCGSKVKGKQDNQNFILTAQASPSNTVTKIAS 130
 DB 74 HTKI---RHVDAHATLSD-GYVVQVVMGLLSNGSPERKFMOTFVLPAEGSVPNKEY-VHN 128
 QY 131 DCFRFOD 137
 DB 129 DMFRIED 135

RESULT 12
 G3B2_MOUSE
 ID G3B2_MOUSE STANDARD; PRT; 482 AA.
 AC P97379; Q9R1B8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ras-GTPase-activating protein binding protein 2 (GAP SH3-domain binding protein 2) (G3BP-2).
 GN G3BP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX -NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RX MEDLINE-98236259; PubMed-9575347;
 RA Kennedy D., Wood S.A., Ramsdale T., Tam P.P., Steiner K.A., Mattick J.S.;
 RT "Identification of a mouse orthologue of the human ras-GAP-SH3-domain binding protein and structural confirmation that these proteins contain an RNA recognition motif.";
 RL Biomed. Pept. Proteins Nucleic Acids 2:93-99(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
 RA Kennedy D., Mattick J.S.;
 RT "Characterisation and chromosomal location of G3BP-1 and G3BP-2a/b, members of a novel SH3 domain-binding and RNA-binding protein family implicated in signal transduction.";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA TRANSPORT (POTENTIAL).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 DR EMBL; U65313; AAC53553.1; -
 DR EMBL; AF145285; AAD51933.1; -
 DR HSSP; P09651; 1HA1.
 DR MGD; MGI:1346339; G3bp2.
 DR InterPro; IPR002075; NTF2.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; PF02136; NTF2; 1.
 DR PROSITE; PS00360; RRM; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS0177; NTF2_DOMAIN; 1.
 DR PROSITE; PS0102; RRM_RNP_1; FALSE_NEG.
 KW Transport; RNA-binding; Alternative splicing.
 FT DOMAIN 11 133 NTF2.
 FT DOMAIN 331 409 RNA-BINDING (RRM).
 FT DOMAIN 134 223 GLU-RICH.
 FT VARSPPLIC 243 275 MISSING (IN ISOFORM B).
 FT CONFLICT 267 267 S -> P (IN REF. 2).
 FT CONFLICT 359 359 E -> V (IN REF. 3).
 FT CONFLICT 460 460 M -> I (IN REF. 3).
 SQ SEQUENCE 482 AA; 54111 MW; 2545CGA3FAAE218 CRC64;

FT CONFLICT 476 482 RETGRR -> TLHRTASLKSCHWQFWQWYLTHRVCILVKE
 FT FVLNVTOPF (IN REF. 1).
 SQ SEQUENCE 482 AA; 54088 MW; FB8BC2086123A5CE CRC64;

Query Match 15.8%; Score 115.5; DB 1; Length 482;
 Best Local Similarity 29.9%; Pred. No. 0.00021;
 Matches 38; Conservative 18; Mismatches 58; Indels 13; Gaps 4;

QY 19 EFVNYVTMDKRRRLSLRYNGTATLVNG-----NAVSGQSLSEFFEMLPSSF 70
 DB 14 EFVRYTYLLNKAPEYLHRYGRNSYVGGVDASKPOENVYGVNDIHKVLSLNFSEC 73
 QY 71 QISVVDQCPVDEATPSQTTVLVVICGSKVFEKNQDFNQNFILTAQASPSNTVWKIAS 130
 DB 74 HKKI---RHVDAHATLSD-GVVYVVGMLLSNSGQPKFEMQFVLAPGSPVKNFY-VHN 128
 QY 131 DCFRFQD 137
 DB 129 DMFRYED 135

RESULT 13
 NTF2_ORYSA STANDARD; PRT; 122 AA.
 AC Q9XJ54;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Nuclear transport factor 2 (NTF-2).
 GN NTF2.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cy. Nipponbare;
 RA Matsukura R., Iwasaki T., Jiang C., Yamamoto N.;
 RT "Molecular cloning of a cDNA encoding nuclear transport factor 2
 (NTF2) from rice."
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE
 PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
 AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; AB011262; BAA81910.1;
 DR HSP; P13662; IAK2.
 DR InterPro: IPR002075; NTF2.
 DR Pfam: PF02136; NTF2; 1.
 DR PROSITE; PS50177; NTF2_DOMAIN; 1.
 KW Transport; Protein transport.
 FT DOMAIN 6 119 NTF2.
 SQ SEQUENCE 122 AA; 13354 MW; 84100EBB62AB1BD1 CRC64;

Query Match 14.6%; Score 107; DB 1; Length 122;
 Best Local Similarity 28.9%; Pred. No. 0.0003;
 Matches 35; Conservative 21; Mismatches 53; Indels 12; Gaps 4;

QY 17 AEFVNYVTMDKRRRLSLRYNGTATLVNGNAVSGQSLSEFFEMLPSSF--FOISV 74
 DB 7 AKAFVEHYRTFTDNRPALVSLYDQSGMLTFEGQOFLGAAAIAGKGLSLPFAQCHDINT 66

QY 75 VDCQPVDEATPSQTTVLVVICGSKVFEKNQ-RDFNQNFILTAQASPSNTVWKIASDCF 133
 DB 67 VDCQP-----SGPQGMVLFVSSSLRTGPDHPLKFSOMF-----QLLPAGGNFYQNDMF 117

QY 134 R 134
 DB 118 R 118

RESULT 14
 NTF2_SCHPO STANDARD; PRT; 123 AA.
 ID NTF2_SCHPO
 AC Q10100; Q9UTF4;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Probable nuclear transport factor 2 (NTF-2).
 GN NTF2 OR SPAC1B9.01C OR SPAC15F9.03C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE-21848401; PubMed-11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Weltyens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Coffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE
 PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
 AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; AL109951; CAB53052.1;
 DR EMBL; Z68197; CAA92380.2;
 DR HSP; P13662; IAK2.
 DR InterPro: IPR002075; NTF2.
 DR Pfam: PF02136; NTF2; 1.
 DR PROSITE; PS50177; NTF2_DOMAIN; 1.

KW Transport; Protein transport.
 FT DOMAIN 7 120 NTF2.
 FT CONFLICT 75 76 TT-> SL (IN REF. 1; CAA92380).
 SQ SEQUENCE 123 AA; 14078 MW; F387303CAA3270E6 CRC64;

Query Match 14.6%; Score 107; DB 1; Length 123;
 Best Local Similarity 28.9%; Pred. No. 0.0003;
 Matches 35; Conservative 21; Mismatches 53; Indels 12; Gaps 4;

OY 17 AEEFNVYTTMDKRRRLSLRYNGTATLVNNGNAVSGOESLSEFFEMLPSEFO--ISV 74
 DB 8 ATQCFQFYQTDFSDRSQSLSSLYREESMLSFEGALQGTKAIVERKLVSILPFORVQHRIST 67
 OY 75 VDCQPVDHEATPSQTTVLVVICGSKVKEGKQKQDFNQNFIILTAQSPSNTVWKIASDCF 133
 DB 68 LDAQP-----TGTCGSIIVWTGELLDEEQAQRIYSQVHLV-----NNNGNYIVLNDLF 118
 OY 134 R 134
 DB 119 R 119

RESULT 15
 G3BP_HUMAN STANDARD; PRT; 466 AA.
 AC Q13283;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ras-GTPase-activating protein binding protein 1 (GAP SH3-domain
 binding protein 1) (G3BP-1).
 GN G3BP OR G3BP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=96220439; PubMed=8649363;
 RA Parker F., Maurier F., Delumeau I., Duchesne M., Faucher D.,
 RA Debussche L., Duque A., Schweighoffer F., Tocque B.;
 RT "A Ras-GTPase-activating protein SH3-domain-binding protein.";
 RL Mol. Cell. Biol. 16:2561-2569(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC Sub-UNIT: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA
 CC TRANSPORT (POTENTIAL).
 CC Sub-UNIT: BINDS TO THE SH3 DOMAIN OF RAS-GTPASE-ACTIVATING PROTEIN
 CC (GAP120).
 CC Sub-CELLULAR LOCATION: Cytoplasmic (Probable).
 CC TISSUE SPECIFICITY: UBIQUITOUS.
 CC SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
 CC SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; U32519; AAB07787.1; -
 DR EMBL; BC006997; AAB06997.1; -
 DR HSSP; P09651; 1HA1.
 DR InterPro; IPR002075; NTF2.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR Pfam; PF02136; NTF2; 1.

DR SMART: SMO0360; RRM; 1.
 DR PROSITE; PS0177; NTF2_DOMAIN; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS0030; RRM_RNP_1; FALSE_NEG.
 KW Transport; RNA-binding.
 FT DOMAIN 11 133 NTF2
 FT DOMAIN 340 415 RNA-BINDING (RRM).
 FT DOMAIN 144 225 GLU-RICH.
 FT DOMAIN 430 461 GLY-RICH.
 SQ SEQUENCE 466 AA; 52164 MW; 0F9429D78E0C7F59 CRC64;

Query Match 14.4%; Score 105.5; DB 1; Length 466;
 Best Local Similarity 28.3%; Pred. No. 0.0022;
 Matches 36; Conservative 21; Mismatches 57; Indels 13; Gaps 4;

OY 19 EFNVVYTTMDKRRRLSLRYNGTATLVNNG-----NAVSGOESLSEFFEMLPSEF 70
 DB 14 EFNQYTYLLNQAPDMLHRFYCKNSYVHGGLDSNGKPADAVYGK---EHRKVMNQNF 70
 OY 71 QISVVDCQPVHDEATPSQTTVLVVICGSKVKEGKQKQDFNQNFIILTAQSPSNTVWKIAS 130
 DB 71 TNCHTKIRHYDAHATLND-GYVVQVMGLLSNNQALRRFMOTFVLAPEGSVANKEY-VHN 128
 OY 131 DCFRFOD 137
 DB 129 DIFRYQD 135

Search completed: March 4, 2003, 14:59:38
 Job time : 4.35122 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:32 ; Search time 9.59142 Seconds

(without alignments)
3007.543 Million cell updates/sec

Title: US-09-763-902b-2

Perfect score: 732

Sequence: 1 MASVDFTYVDQACRAAEF.....PSNTWKIASDCFRFQDNAS 140

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_undefined.*
- 15: sp_rvirus.*
- 16: sp_bacteriopl.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151.5	20.7	167	5	Q8T2G7 dictyosteli
2	142	19.4	200	3	Q8HBY6 schizosacch
3	135	18.4	146	10	Q8RU32 oryza sativ
4	128.5	17.6	127	5	Q8T2D5 dictyosteli
5	125.5	17.1	123	3	Q9P8H0 yarrowia li
6	119	16.3	688	5	Q9I7J5 drosophila
7	119	16.3	690	5	Q9NH72 drosophila
8	119	16.3	690	5	Q9VFT4 drosophila
9	116.5	15.9	522	10	Q9LPY0 arabidopsis
10	112.5	15.4	125	3	Q96VN3 emericella
11	106	14.5	537	5	Q21351 caenorhabdi
12	105.5	14.4	130	5	Q9VJ85 drosophila
13	105	14.3	122	10	Q9FZK4 arabidopsis
14	91.5	12.5	450	10	Q9FND0 arabidopsis
15	88.5	12.1	130	5	Q9VRD6 drosophila
16	87.5	12.0	460	10	Q9FME2 arabidopsis

17	84	11.5	946	10	Q9SFV5	Q9sfv5 arabidopsis
18	83	11.3	427	10	Q9LDI9	Q9ldi9 arabidopsis
19	80.5	11.0	428	10	Q9LMD6	Q9lmd6 arabidopsis
20	77.5	10.6	398	11	Q07763	Q07763 mus musculu
21	76.5	10.5	1211	10	Q9ATY7	Q9aty7 linum usita
22	74.5	10.2	397	11	Q9JIE0	Q9jie0 mus musculu
23	74.5	10.2	1939	13	Q9PTY2	Q9pty2 gallus gall
24	74	10.1	637	10	Q9SF50	Q9sf50 arabidopsis
25	74	10.1	1044	2	Q54354	Q54354 anabaena sp
26	74	10.1	1044	16	Q8YTA6	Q8yta6 anabaena sp
27	73.5	10.0	666	3	Q8TFR8	Q8tfr8 saccharomyc
28	73.5	10.0	666	3	Q8TFR7	Q8tfr7 saccharomyc
29	73.5	10.0	685	3	Q9HGV1	Q9hgv1 saccharomyc
30	73.5	10.0	685	3	Q8TFB8	Q8tfb8 saccharomyc
31	72.5	9.9	452	3	Q9UTG3	Q9utg3 schizosacch
32	72.5	9.9	488	10	Q9SAA9	Q9saa9 arabidopsis
33	72	9.8	178	5	Q9VGQ4	Q9vgq4 drosophila
34	72	9.8	261	5	Q9VGR3	Q9vgr3 drosophila
35	72	9.8	262	5	Q95SH9	Q95sh9 drosophila
36	72	9.8	262	5	Q8T3W8	Q8t3w8 drosophila
37	72	9.8	459	5	Q44024	Q44024 reticulomyc
38	71.5	9.8	473	10	Q9LSG2	Q9lsq2 arabidopsis
39	71.5	9.8	488	10	Q8VIJ4	Q8vyj4 arabidopsis
40	71.5	9.8	492	17	Q8TNS3	Q8tns3 methanosarc
41	71.5	9.8	550	12	Q9IV37	Q9iv37 human calic
42	71.5	9.8	550	12	Q918B6	Q918b6 norwalk-vlr
43	71.5	9.8	550	12	Q8V778	Q8v778 norwalk-vlr
44	71.5	9.8	747	3	Q8TFX3	Q8tfx3 aspergillus
45	71.5	9.8	895	2	Q52507	Q52507 serratia li

ALIGNMENTS

RESULT 1

Q8T2G7	PRELIMINARY;	PRT;	167	AA.
ID	Q8T2G7			
AC	Q8T2G7;			
DT	01-JUN-2002 (TREMREL. 21, Created)			
DT	01-JUN-2002 (TREMREL. 21, Last sequence update)			
DT	01-JUN-2002 (TREMREL. 21, Last annotation update)			
DE	Hypothetical 19.0 kDa protein.			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.			
OX	NCBI_TaxID=44689;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AX4;			
RA	Gloekner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,			
RA	Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,			
RA	Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;			
RT	"Sequence and Analysis of Chromosome 2 of Dictyostelium."			
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC115607; AAL92371.1;			
KW	Hypothetical protein.			
SQ	SEQUENCE 167 AA; E16E4A5F4DA00BAA CRC64;			

Query Match	20.7%	Score 151.5;	DB 5;	Length 167;
Best Local Similarity	29.0%	Pred. No. 1e-08;		
Matches	38;	Conservative	22;	Mismatches 64; Indels 7; Gaps 4;

Qy	7	KTYVDQACRAAEFV-NVYTTMDKRELLSRLYMGATLVNNGNAVSGQSLSEFFEML	65
Db	38	KKIYETITPRAEFYKPYKDYSSRADLGLYKDHDSVSWNGTECKGPBHIKLLAEI	97
Qy	66	PSSEFQISVDCQPV--HDEATPSQTTVLVVICGSKPEGNKQKDFNQNFILTAQSPSN	123
Db	98	PNSHVHVFETPAQVPSPDDKENPN---ILITATGKVTYKTTTSQHOFHTFLLVKPTNSN	154
Qy	124	TWKIASDCFR 134	
Db	155	LFYLSYDCIR 164	

RESULT 2
 Q9HDY6 PRELIMINARY; PRT; 200 AA.
 AC Q9HDY6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative nuclear export factor.
 GN SPAPB1A10.03
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA O'Neill S., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL512562; CAC21476.1;
 DR InterPro; IPR02075; NTF2.
 DR Pfam; PF02136; NTF2; 1.
 SQ SEQUENCE 200 AA; 23029 MW; B05E47D62B27B7FF CRC64;

Query Match 19.4%; Score 142; DB 3; Length 200;
 Best Local Similarity 25.6%; Pred. No. 1.4e-07;
 Matches 31; Conservative 33; Mismatches 43; Indels 14; Gaps 4;
 QY 10 VDOACRAAEFFVNYVTMDKRRLLSRLYMGATLVWNGNAVSGESLSEFFEM---LP 66
 DB 1 MESSVYKAEQFVQRYSSLDNRNGAERYNSLLWNGKPM---QVTEFTSLVNL 56

QY 67 SFEQISVDCOPVHDEATPSQTTVLVWICGVKFEKGNQKQDFNQNFILTAQASPSNTVW 126
 DB 57 YSKTKVEDPDSQV-----MGNDNIIIVSGTIREDKPKPHVFSYR--SATTQDDTTF 109
 QY 127 K 127
 DB 110 K 110

RESULT 3
 Q8RU32 PRELIMINARY; PRT; 146 AA.
 AC Q8RU32;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative nuclear transport factor 2.
 GN P0415A04.24.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC clone:P0415A04."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003345; BAB90110.1;
 SQ SEQUENCE 146 AA; 15796 MW; 496BE9A02E1E9670 CRC64;

Query Match 18.4%; Score 135; DB 10; Length 146;
 Best Local Similarity 32.8%; Pred. No. 5.7e-07;
 Matches 42; Conservative 19; Mismatches 55; Indels 12; Gaps 5;
 QY 11 DOACRAAEFFVNYVTMDKRRLLSRLYMGATLVWNGNAVSGESLSEFFEMLPSPSEF 70
 DB 23 DQCDVVARAFVEYVYVTFDTNRAALAAALYQTSLSLSEFEGHNVAGAEETGRKLLGLPFFQC 82

QY 71 Q--ISVDCOPVHDEATPS-QTTVLVWICGVKFEKGNK-QRDFNQNFILTAQASPSNTVW 126
 DB 83 RHAVCTVDCQP-----TPSPGGILFVSGNLQALAGEHOLRFSONFQVLVPNEQGS---F 134

QY 127 KIAADCFR 134
 DB 135 FVQNDIFR 142

RESULT 4
 Q8T2D5 PRELIMINARY; PRT; 127 AA.
 AC Q8T2D5;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative nuclear transport factor similar to nuclear transport factor 2 (NTF2) from.
 DE 2 (NTF2) from.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RT Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC115613; AAM10776.1;
 SQ SEQUENCE 127 AA; 14317 MW; 2D45B1B20021B298 CRC64;

Query Match 17.6%; Score 128.5; DB 5; Length 127;
 Best Local Similarity 25.5%; Pred. No. 2.5e-06;
 Matches 35; Conservative 29; Mismatches 56; Indels 17; Gaps 5;
 QY 1 MASVDFKTYVDOACRAAEFFVNYVTMDKRRLLSRLYMGATLVWNGNAVSGESLSE 60
 DB 1 MOSVD-----POVGVGKQFVEHYGIFDSNRAGLTQYQQQTTLTWEGKFLSGADATVK 55

QY 61 FFEMLP--SSEFQISVDCOPVHDEATPSQTTVLVWICGVKFEKGNQKQDFNQ--NFILTA 117
 DB 56 HIVELPFOOTNRKINSIDCQTY-----QPGIMVTGTLLIDGEAK---NOLKFEVQVF 106

QY 118 QASPSNTVWKIASDCFR 134
 DB 107 NLASNGSFLINDFFR 123

RESULT 5.
 Q9P8H0 PRELIMINARY; PRT; 123 AA.
 AC Q9P8H0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Rph1.
 GN RPH1.
 OS Yarrowia lipolytica (Candida lipolytica).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Dipodascaceae; Yarrowia.
 OX NCBI_TaxID=4952;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Cordero Otero R.R., Lepingle A., Gaillardin C.;
 RA "RPH1 gene from Yarrowia lipolytica."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rojas Quijano R., Lepingle A., Gaillardin C.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF260231; AAF70316.1;

DR HSP: p13662; 100N.
 DR InterPro: IPR002075; NTF2.
 DR Pfam: PF02136; NTF2; 1.
 SQ SEQUENCE 123 AA; 13696 MW; 386CB95CF0058251 CRC64;

Query Match 17.1%; Score 125.5; DB 3; Length 123;
 Best Local Similarity 29.9%; Pred. No. 5e-06;
 Matches 40; Conservative 20; Mismatches 57; Indels 17; Gaps 5;
 QY 3 SVDFYTYVDQACRAAEFVNYYTMDKRRLLSLRYMGATLVWNGNAVSGQESLSEFF 62
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2 SVDFNTLAKQFC---EF---YQFTFDRSOLGNLYRDHSMLTGTGQHQGAQAIKVL 54
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 63 EMLPSE--FQISVVDQCPVHDEATPSQTTVLVVICGSKVEGKNQKQDFNQNFILTAQAS 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 55 VGLPGQGVNKHSDIAQF-----ASAQGDVILVIVGELCVGDNDPLPGQVFHLI---- 106
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 PSNTYWKIASDCFR 134

Db 107 PDGSSVYVFNDFR 120

RESULT 6

Q917J5 PRELIMINARY; PRT; 688 AA.
 AC Q917J5;
 DT 01-WAR-2001 (TRENBLrel. 16, Created)
 DT 01-WAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Rin protein.
 GN RIN OR CG9412.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Efannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; "
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003701; AAG22151.1;
 DR FlyBase; FBgn0015778; rin.
 DR InterPro; IPR002075; NTF2.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF02136; NTF2; 1.
 DR Pfam; PF00076; rin; 1.
 DR SMART; SM00360; rrm; 1.
 DR PROSITE; PS50102; RRM; 1.
 SQ SEQUENCE 688 AA; 74713 MW; 9E329949E7248C19 CRC64;

Query Match 16.3%; Score 119; DB 5; Length 688;
 Best Local Similarity 30.3%; Pred. No. 0.00023;
 Matches 37; Conservative 19; Mismatches 54; Indels 12; Gaps 5;

QY 19 EFNYYVTMDKRRLLSLRYMGATLVWNGNAVSGQESLSEFFEMLPSE--FOISV 75

Db 17 EFVROYTLLNKAPNHLHRYFNHNSSYIHGESKLVVGQREIHNRIQQLNFNDCHAKISQV 76

QY 76 DCQPVHDEATPSQTTVLVVICGSKVEGKNQKQDFNQNFILTAQASPSNTVWKIASDCFR 135

Db 77 DAQATLNG-----VVQVVTGELSDNGQPMRRFTQTFVLAQ-SPKK--YVVDHDIERY 127

QY 136 QD 137

Db 128 QD 129

RESULT 7

Q9NH72 PRELIMINARY; PRT; 690 AA.
 AC Q9NH72;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Rasputin.
 GN RIN OR CG9412.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20191744; PubMed=10725247;
 RA Pazman C., Mayes C.A., Fanto M., Haynes S.R., Mlodzik M.,
 RT Rasputin, the Drosophila homolog of the RasGAP SH3 binding protein,
 RT functions in Ras and Rho mediated signaling.";
 RL Development 127:1715-1725(2000).
 DR EMBL; AF231031; AAF68949.1;
 DR FlyBase; FBgn0015778; rin.
 DR InterPro; IPR002075; NTF2.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF02136; NTF2; 1.
 DR Pfam; PF00076; rin; 1.
 DR SMART; SM00360; rrm; 1.
 DR PROSITE; PS50102; RRM; 1.
 SQ SEQUENCE 690 AA; 74985 MW; 40C5BD7E9F5FA253 CRC64;

Query Match 16.3%; Score 119; DB 5; Length 690;
 Best Local Similarity 30.3%; Pred. No. 0.00023;
 Matches 37; Conservative 19; Mismatches 54; Indels 12; Gaps 5;

QY 19 EFNYYVTMDKRRLLSLRYMGATLVWNGNAVSGQESLSEFFEMLPSE--FOISV 75

Db 19 EFVROYTLLNKAPNHLHRYFNHNSSYIHGESKLVVGQREIHNRIQQLNFNDCHAKISQV 78

QY 76 DCQPVHDEATPSQTTVLVVICGSKVEGKNQKQDFNQNFILTAQASPSNTVWKIASDCFR 135


```

ID: Q96VN3 PRELIMINARY; PRT; 125 AA.
AC Q96VN3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Nuclear transport factor 2.
DE NTF2.
GN Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA Espeso E.A., Penalva M.A.;
RT "Nuclear transport factor 2 from Aspergillus nidulans.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY038983; AAK71467.1;
DR InterPro; IPR002075; NTF2.
DR Pfam; PF021136; NTF2; 1.
SQ SEQUENCE 125 AA; 13871 MW; 8FC18BDFC597E2F2 CRC64;

Query Match 15.4%; Score 112.5; DB 3; Length 125;
Best Local Similarity 27.3%; Pred. No. 0.00014;
Matches 33; Conservative 24; Mismatches 53; Indels 11; Gaps 4;

QY 17 ABEFVNYVTMDKRRLLSLRYMGATLVWNGNAVSGQESLSSEFFEMLP--SSEFOISV 74
Db 8 AQOFVTFYQTFDGNRAGLALFRDHSMLTFTSAQGVAGIIEKLTSLPQKVKHQVST 67

QY 75 VDCQPVHDEATPSQTTVLVVICGSVFEGNKQ-RDFNQNFILTAQAQSPSNVTKIASDCF 133
Db 68 LDAQPSGEHG-----GLVLVTGALLVDEKNPMNYTQTFQLMPDGAGS---YFVLNDVF 119

QY 134 R 134
Db 120 R 120

RESULT 11
Q21351 PRELIMINARY; PRT; 537 AA.
ID Q21351
AC Q21351;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE K08F4.2 protein.
GN K08F4.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Hembry C.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL Science 282:2012-2018(1998).
DR EMBL; Z68879; CAA93082.1;
DR InterPro; IPR002075; NTF2.
DR Pfam; PF021136; NTF2; 1.
DR Pfam; PF000076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR SMART; PS01012; RRM; 1.
SQ SEQUENCE 537 AA; 59730 MW; 7C3169BFA71C312F CRC64;

Query Match 14.5%; Score 106; DB 5; Length 537;
Best Local Similarity 23.6%; Pred. No. 0.0046;

```

```

Matches 30; Conservative 31; Mismatches 48; Indels 18; Gaps 5;

QY 16 AAE-----FVNYVTMDKRRLLSLRYMGATLVWNGNAVSGQESLSSEFFEMLP--SSEF 70
Db 96 AAEQVGGAFCFOFIYTVSENKRAITFRYGHESKFYDDQVVTGSQEIYKLYNLPTTHF 155

QY 71 QISVDCQPVHDEATPSQTTVLVVICGSVFEGNKQNFILTAQAQSPSNVTKIAS 130
Db 156 KIQSIKGP-----TPKQGVIIINVIGTVNL-----RPFQSELLGQGGKK---YVES 202

QY 131 DCPRFOD 137
Db 203 DAFOYID 209

RESULT 12
Q9VJ85 PRELIMINARY; PRT; 130 AA.
ID Q9VJ85
AC Q9VJ85;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG10174 protein.
GN CG10174.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003659; AAF53669.1;
DR HSSP; P13662; IOUN.

```

```
DR FlyBase; FBgn0032680; CG10174.
DR InterPro; IPR002075; NTF2.
DR Pfam; PF02136; NTF2; 1.
-SQ SEQUENCE 130 AA; 14582 MW; 39286FBC5BD7C34E CRC64;

Query Match 14.48; Score 105.5; DB 5; Length 130;
Best Local Similarity 28.08; Pred. No. 0.00085;
Matches 35; Conservative 19; Mismatches 50; Indels 21; Gaps 6;

Qy 18 EEFVNVYITMD--KRRLLSLRYMGT-ATLVWNGNAVSGESLSEFFEMLPSSFEQ--- 71
Db 12 KEFVQYVIAFDPPANREVINFNATDSFTFEGNOIQGAPKI---LEKVQSLSFQKIA 68
Qy 72 --ISVDCQPHDEATPSQTTVLVWNGNAVSGESLSEFFEMLPSSFEQ--- 129
Db 69 RVITVDSQPTSDGG-----VLIIVLGRKLCDDPPHAFSQIFLL-----KPNGGSLFVA 118
Qy 130 SDCFR 134
Db 119 HDIFR 123

RESULT 13
Q9FZK4 PRELIMINARY; PRT; 122 AA.
AC Q9FZK4
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE F17L21.10 (Similar to nuclear transport factor 2).
GN F17L21.10
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson-Hopson C., Brooks S., Buehler E., Chao Q., Khan S., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chioi J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F17L21 from chromosome
RT I.";
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chioi J., Choi E.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;

Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[7]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
[8]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004557; AAF99749.1;
DR EMBL; AY042889; AAL68829.1;
DR EMBL; AY072473; AAL68888.1;
DR HSP; P13662; IOUN.
DR InterPro; IPR002075; NTF2.
DR Pfam; PF02136; NTF2; 1.
SQ SEQUENCE 122 AA; 13527 MW; E7CDD9486631A1D2 CRC64;

Query Match 14.38; Score 105; DB 10; Length 122;
Best Local Similarity 27.38; Pred. No. 0.00089;
Matches 33; Conservative 22; Mismatches 54; Indels 12; Gaps 4;

Qy 17 AEEFNVYITMDKRRLLSLRYMGTATLVWNGNAVSGESLSEFFEMLP--SSEFQISV 74
Db 7 AKAEVHYITFDANRPGVLVSGYQEGSLTFEGKIQGQSNIVAKLTGLPQCKKHMT 66
Qy 75 VDCQPHDEATPSQTTVLVWNGNAVSGESLSEFFEMLP--SSEFQISV 133
Db 67 VDCQPHDEATPSQTTVLVWNGNAVSGESLSEFFEMLP--SSEFQISV 117
Qy 134 R 134
Db 118 R 118

RESULT 14
Q9FND0 PRELIMINARY; PRT; 450 AA.
AC Q9FND0
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE GBIAD20086.1 (Hypothetical 50.1 kDa protein).
GN AT5G43960.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
```

Search completed: March 4, 2003, 15:02:32
Job time : 12.5914 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:55:07 ; Search time 4.85349 Seconds
(without alignments)
848.711 Million cell updates/sec

Title: US-09-763-902B-2
Perfect score: 732
Sequence: 1 MASVDFTYVDQACRAEEF.....PSNTWKIASDCFRQDNAS 140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2.6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep:*

*Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	12.3	466	2	US-08-836-791-9
2	68.5	9.4	519	4	US-09-172-841-55
3	68.5	9.4	2628	2	US-08-570-311-14
4	66.5	9.1	569	1	US-07-821-716-2
5	66.5	9.1	569	2	US-08-381-603-2
6	66.5	9.1	569	4	US-08-924-376-2
7	66.5	9.1	569	4	US-08-685-212-2
8	66.5	9.1	569	4	US-09-173-151A-31
9	66.5	9.1	569	4	US-08-466-332A-2
10	66.5	9.1	569	5	PCT-US94-02414-2
11	66.5	9.1	569	5	PCT-US96-08899-2
12	65.5	8.9	1871	2	US-08-694-869-1
13	65.5	8.9	1871	3	US-09-349-546-1
14	65	8.9	329	4	US-09-651-200-18
15	64	8.7	396	3	US-09-082-310-2
16	64	8.7	396	4	US-09-575-205-2
17	63	8.6	490	3	US-09-226-741-1
18	63	8.6	490	4	US-09-595-514-1
19	62.5	8.5	310	2	US-08-943-600A-3
20	62.5	8.5	956	1	US-08-185-232A-2
21	62.5	8.5	956	1	US-08-416-523-2
22	62.5	8.5	956	3	US-08-789-478-2
23	62	8.5	766	2	US-08-846-762-5
24	62	8.5	2471	3	US-09-112-450-4
25	62	8.5	2471	4	US-09-419-291A-4
26	61.5	8.4	1698	4	US-09-315-793-12
27	61	8.3	219	3	US-09-078-317-10
					Sequence 9, Appl
					Sequence 55, Appl
					Sequence 14, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 31, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 18, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 1, Appl
					Sequence 3, Appl
					Sequence 2, Appl
					Sequence 5, Appl
					Sequence 4, Appl
					Sequence 12, Appl
					Sequence 10, Appl

28	61	8.3	219	4	US-09-454-818-10	Sequence 10, Appl
29	60.5	8.3	496	1	US-08-313-075A-50	Sequence 50, Appl
30	60.5	8.3	522	4	US-08-961-083-120	Sequence 120, Appl
31	60.5	8.3	1040	4	US-08-961-083-118	Sequence 118, Appl
32	60	8.2	897	1	US-08-095-737-4	Sequence 4, Appl
33	60	8.2	897	1	US-08-480-145-4	Sequence 4, Appl
34	60	8.2	897	2	US-08-477-389-4	Sequence 4, Appl
35	60	8.2	998	2	US-08-449-645A-20	Sequence 20, Appl
36	60	8.2	998	2	US-08-702-367A-20	Sequence 20, Appl
37	60	8.2	998	5	PCT-US95-04681-20	Sequence 20, Appl
38	60	8.2	1732	2	US-08-570-311-10	Sequence 10, Appl
39	60	8.2	1732	2	US-08-353-485-10	Sequence 10, Appl
40	59.5	8.1	310	2	US-08-943-600A-1	Sequence 1, Appl
41	59.5	8.1	863	4	US-09-619-353-14	Sequence 14, Appl
42	59.5	8.1	1335	4	US-09-134-001C-3716	Sequence 3716, Appl
43	59.5	8.1	2386	2	US-09-016-366A-12	Sequence 12, Appl
44	59	8.1	193	3	US-08-655-352-2	Sequence 2, Appl
45	59	8.1	193	4	US-09-258-016-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-836-791-9
; Sequence 9, Application US/08836791
; Patent No. 5886150
; GENERAL INFORMATION:
; APPLICANT: Duchesne, Marc
; APPLICANT: Faucher, Didier
; APPLICANT: Parker, Fabienne
; APPLICANT: Schweighoffer, Fabien
; APPLICANT: Toque, Bruno
; TITLE OF INVENTION: Peptides Capable of Binding to the GAP
; TITLE OF INVENTION: Protein SH3 Domain, Nucleotide Sequences Coding Therefor,
; TITLE OF INVENTION: and Preparation and Use Thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, Mailstop 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,791
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94/13955
; FILING DATE: 22-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 95/05753
; FILING DATE: 16-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR95/01539
; FILING DATE: 22-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fehliner Esq., Paul F.
; REGISTRATION NUMBER: 35,135
; REFERENCE/DOCKET NUMBER: ST9408361-US
; TELEPHONE: (610) 454-3839
; TELEFAX: (610) 454-3808
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466 amino acids
; TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-791-9

Query Match 12.3%; Score 90; DB 2; Length 466;
Best Local Similarity 27.6%; Pred. No. 0.011;
Matches 35; Conservative 21; Mismatches 57; Indels 14; Gaps 5;

QY 19 EFNVTYTTMDKRRRLSLRYMGATLVNG-----NAVSGQSLSEFFEMLPSSSEF 70
DB 14 EFRQYITLNPADMDHFRFGKNSYVGGGLDSNGKPADAVYGQR---EIHRYVMSQNF 70
QY 71 QISVDCQPVHDEATQNTLVVVICGSKFEKGNKQKQDNQNFILTAQASPSNTVVKIAS 130
DB 71 TNCHTKIRHVDAHATLND-GVVYVNGVLLNNQALRRFNQFVLAPESVANKFY-VHN 128

QY 131 DCFRFOD 137

DB 129 D-IRYOD 134

RESULT 2

US-09-172-841-55
Sequence 55, Application US/09172841
Patent No. 6232081

GENERAL INFORMATION:
APPLICANT: Harper, Jeffrey W.
APPLICANT: Elledge, Stephen J.
TITLE OF INVENTION: F-BOX PROTEINS AND GENES
FILE REFERENCE: BCM-03510
CURRENT APPLICATION NUMBER: US/09/172,841
EARLIER FILING DATE: 1998-10-15
EARLIER APPLICATION NUMBER: 08/951,621
NUMBER OF SEQ ID NOS: 60
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 55
LENGTH: 519
TYPE: PRT
ORGANISM: Homo sapiens
US-09-172-841-55

Query Match 9.4%; Score 68.5; DB 4; Length 519;
Best Local Similarity 19.7%; Pred. No. 5.9;
Matches 30; Conservative 27; Mismatches 48; Indels 47; Gaps 6;

QY 10 VDOACRAAEFVNVTYTTMDKRRRLSLR-----YMGTA--TLVWNGNAVSGQSLSEF 61
DB 122 IDESESAEESIAISIAQMEK--RLHGLHNVLPVGTGTVKTLVLAYSNAVSSKMWVQI 179
QY 62 FEMLPSSSEF---QISVVD-----COPVHDEATPSQTTVLVVI 95
DB 180 LELCPNLEHLDITQDIDSAPDSWGLGCCOSLRHLDLSCGKITDVALEKISRALGIL 239
QY 96 CGSVKFEKGNKQKQDNQNFILTAQASPSNTVVK 127
DB 240 -----TSHQSGFLKTSKTSKITSTANK 260

RESULT 3

US-08-570-311-14;
Sequence 14, Application US/08570311
Patent No. 5824791

GENERAL INFORMATION:
APPLICANT: Progulskis-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guyvaline
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
AND PROBES FOR THE DETECTION OF PERIODONTAL DISEASE

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/353,485

FILING DATE: 09-DEC-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/647,119

FILING DATE: 25-JAN-1991

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/241,640

FILING DATE: 08-SEP-1988

ATTORNEY/AGENT INFORMATION:

NAME: Whitlock, Ted W.

REGISTRATION NUMBER: 36,965

REFERENCE/DOCKET NUMBER: UFI5.C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 2628 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-570-311-14

Query Match 9.4%; Score 68.5; DB 2; Length 2628;
Best Local Similarity 27.6%; Pred. No. 61;
Matches 43; Conservative 15; Mismatches 53; Indels 45; Gaps 11;

QY 2 ASVDFKTVVDQACRAAEFVNVTYTTMDKRRRLSLRYMGTA-----TLVW---NG--NA 51
DB 456 AGVSPKVKCKDVTVEGNEFAHVQNLT-----GSAVGQKVTLKWDAPNGTNP 502
QY 52 VSGQSLSEFFEM-LPSSEFQISVVDCQPVHDEAT---PSQTTVL---VVICGS---V 99
DB 503 NPGTTTLESSEFNGIPAS---WKTIDAGDGNNTWTTTPPGGTSPAGHNSAICASSAYI 559
QY 100 KFEKNQKQDFNQNFILTAQASPSN---TYWKIASD 131
DB 560 NFEQPQNDP---NVLVTPELSLPNGGTLTFWVCAOD 592

RESULT 4

US-07-821-716-2

Sequence 2, Application US/07821716

Patent No. 5319071

GENERAL INFORMATION:

APPLICANT: Dower, Steven K.

APPLICANT: March, Carl J.

APPLICANT: Sims, John

APPLICANT: Urdal, David L.

TITLE OF INVENTION: Soluble Interleukin-1 Receptors

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Immunex Corporation

APPLICATION NUMBER: PCT/US96/08899
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gould, Jr., Lewis F.
REGISTRATION NUMBER: 25,057
REFERENCE/DOCKET NUMBER: 109070-12A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 575-6000
TELEFAX: (215) 575-6015
TELEX: 866172
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-08899-2

Query Match 9.18; Score 66.5; DB 5; Length 569;
Best Local Similarity 22.58; Pred. No. 12; Mismatches 53; Indels 13; Gaps 5;
Matches 25; Conservative 20;

Qy 18 EEFVNVYTTMDKRRLLRLMGTATLVNGNAVSGQESLSEFEFEMLPSEFQISVVDC 77
Db 434 EDIVEVENVKSRLLIILVRETSFSLGG--SSEEQIA-WYNALVQDGIRKVLLEL 490
Qy 78 QPVHD-EATPSQTTVLVVICSVKFEKGNQKDFNQNFILTAQASPSNTVK 127
Db 491 EKIQDYKMPESIFKQKHCAIRWSG---DFTQ-----GPQSAKTRFWK 532

RESULT 12
US-08-694-869-1
Sequence 1, Application US/08694869
Patent No. 5994123
GENERAL INFORMATION:
APPLICANT: Olszewski, N.
APPLICANT: Tzafrir, I.
APPLICANT: Somers, D.A.
APPLICANT: Lockhart, B.
APPLICANT: Torbert, K.
TITLE OF INVENTION: SUGARCANE BACILLIFORM VIRUS
TITLE OF INVENTION: PROMOTER
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Schwegman, Lundberg, Woessner & Kluth, P.A.
STREET: P.O. Box 2938
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/694,869
FILING DATE: 09-AUG-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 600.369US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-373-6900
TELEFAX: 612-339-3061
TELEX:
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1871 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-694-869-1

Query Match 8.98; Score 65.5; DB 2; Length 1871;
Best Local Similarity 22.88; Pred. No. 89;
Matches 28; Conservative 19; Mismatches 49; Indels 27; Gaps 6;

Qy 1 MASVDFKTYVDQ-----ACRAAEFVNYYTMDKRRLLSRL-----YMGATATLV-WNG 49
Db 1181 IASIKANEYLOAEAEILVATSEQEFINRSF--MSKNRLLLEEMKEQGYMGEDTLAHWNK 1238
Qy 50 NAVSGQESLSE---FFEMLPSEFQISVVDCQPVHDEA-----TPSQTTVLVVI 95
Db 1239 NQIKCKIELRNPDLIIKDQPTLLNIQKKEAMRKHDALLERKVIKPSKSPHRTNAFIVE 1298
Qy 96 CGS 98
Db 1299 SGT 1301

RESULT 13
US-09-349-546-1
Sequence 1, Application US/09349546
Patent No. 6093569
GENERAL INFORMATION:
APPLICANT: Olszewski, N.
APPLICANT: Tzafrir, I.
APPLICANT: Somers, D.A.
APPLICANT: Lockhart, B.
APPLICANT: Torbert, K.
TITLE OF INVENTION: Sugarcane bacilliform virus promoter
FILE REFERENCE: 600.369US2
CURRENT FILING DATE: 1999-07-08
EARLIER APPLICATION NUMBER: US/09/349,546
EARLIER FILING DATE: 1996-08-09
EARLIER APPLICATION NUMBER: PCT/IB97/01338
EARLIER FILING DATE: 1997-08-13
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1871
TYPE: PRT
ORGANISM: sugarcane bacilliform virus
US-09-349-546-1

Query Match 8.98; Score 65.5; DB 3; Length 1871;
Best Local Similarity 22.88; Pred. No. 89;
Matches 28; Conservative 19; Mismatches 49; Indels 27; Gaps 6;

Qy 1 MASVDFKTYVDQ-----ACRAAEFVNYYTMDKRRLLSRL-----YMGATATLV-WNG 49
Db 1181 IASIKANEYLOAEAEILVATSEQEFINRSF--MSKNRLLLEEMKEQGYMGEDTLAHWNK 1238
Qy 50 NAVSGQESLSE---FFEMLPSEFQISVVDCQPVHDEA-----TPSQTTVLVVI 95
Db 1239 NQIKCKIELRNPDLIIKDQPTLLNIQKKEAMRKHDALLERKVIKPSKSPHRTNAFIVE 1298
Qy 96 CGS 98
Db 1299 SGT 1301

RESULT 14
US-09-651-200-18
Sequence 18, Application US/09651200
Patent No. 6429303
GENERAL INFORMATION:

APPLICANT: Green et al
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
TITLE OF INVENTION: Polypeptides Encoded Thereby
FILE REFERENCE: 15966-562 (CURA-62)
CURRENT APPLICATION NUMBER: US/09/651,200
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/152383
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/172909
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/183578
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18
LENGTH: 329
TYPE: PRT
ORGANISM: Felis catus
US-09-651-200-18

Query Match 8.9%; Score 65; DB 4; Length 329;
Best Local Similarity 27.3%; Pred. No. 8.4;
Matches 30; Conservative 14; Mismatches 42; Indels 24; Gaps 5;

QY 18 EEFVNYVYTTMDKRRLLSRLYMGATLVWNGNAVSGQESLSEF-----PEMLPSSEF 70
Db 178 ENSTKIDYVTKRSQNNVTLYNSISLFF---SVPEAHNVSVFCALKLETLEMLLSLFF 234

QY 71 QISVDCQPVHDEATPSQ-----TTVLVVICGSKVFEGNKQKQDFNQ 111
Db 235 NI---DAQP--NKKDPEQGHFLWIAVLVMEVVFVCGMVSEKTLRKRKKQ 279

RESULT 15

US-09-082-310-2
Sequence 2, Application US/09082310
Patent No. 6096526
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Yue, Henry
TITLE OF INVENTION: HUMAN NUCLEIC ACID METHYLASES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,310
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0520 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: THP1A2S08
CLONE: 2757184
US-09-082-310-2

Query Match 8.7%; Score 64; DB 3; Length 396;
Best Local Similarity 21.3%; Pred. No. 15;
Matches 35; Conservative 17; Mismatches 62; Indels 50; Gaps 5;

QY 2 ASVDKTYVDQACRAAEFEVNYVYTTMDKRRLLSRLYMG-----TATLVWNGNAVSGQES 57
Db 67 ASLDFKRYVTDKRLAETLAQIYLGKPSRPHLLLECNPGPGLTQALLAAGAKVVALES 125

QY 58 LSEFFEMLPSE-----SEFQISVVDC-----QPVHDEATPSQT 89
Db 126 DKTFTPHLESGLKNLDGKLRVHCDFFKLDPRSGGVKPKPAMSSRGLFKNLGIEAVPMTA 185

QY 90 TVLVVICGSKVFEGNKQKQDFNQNFILTAQASPSNTVWKIASDCF 133
Db 186 DIPLKVVGMFSPSRGKR-----ALWKLAYDLY 212

Search completed: March 4, 2003, 15:06:01
Job time: 6.85349 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:58:57 ; Search time 3.35122 Seconds

(without alignments)
1761.695 Million cell updates/sec

Title: US-09-763-902b-2

Perfect score: 732
Sequence: 1 MASVDFKTYVDQACRAAEF.....PSNTWKIASDCFRQDNAS 140

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues 188354

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA.*

- 1: /cgn2.6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
- 2: /cgn2.6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2.6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2.6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2.6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2.6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
- 7: /cgn2.6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
- 8: /cgn2.6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2.6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
- 10: /cgn2.6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*
- 11: /cgn2.6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
- 12: /cgn2.6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
- 13: /cgn2.6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
- 14: /cgn2.6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	732	100.0	164	10	US-09-764-864-1181
2	579	79.1	162	10	US-09-764-864-1180
3	105.5	14.4	477	10	US-09-925-301-977
4	69	9.4	291	10	US-09-864-761-37945
5	68.5	9.4	590	12	US-10-042-417-52
6	67	9.2	632	10	US-09-205-658-159
7	66.5	9.1	569	10	US-09-840-707A-3
8	66.5	9.1	569	10	US-09-731-175-2
9	66.5	9.1	636	10	US-09-205-658-160
10	65	8.9	329	10	US-09-303-510-6
11	65	8.9	329	10	US-09-303-040-6
12	65	8.9	612	10	US-09-758-269-12
13	64	8.7	396	9	US-10-155-613-2
14	64	8.7	407	10	US-09-925-301-1057
15	63.5	8.7	232	9	US-10-058-566-8
16	63	8.6	576	10	US-09-891-216-13
17	63	8.6	684	10	US-09-891-216-14
18	63	8.6	934	10	US-09-891-216-12
19	63	8.6	934	10	US-09-891-216-15

20	62.5	8.5	257	9	US-09-738-626-5977	Sequence 5977, Ap
21	62	8.5	2471	12	US-10-116-048-4	Sequence 4, Appli
22	61.5	8.4	441	9	US-09-738-626-6613	Sequence 6613, Ap
23	61	8.3	1349	10	US-09-815-242-5898	Sequence 5898, Ap
24	61	8.3	1349	10	US-09-815-242-13137	Sequence 13137, A
25	61	8.3	2150	10	US-09-321-987B-2	Sequence 2, Appli
26	61	8.3	2165	10	US-09-800-729-155	Sequence 155, App
27	60.5	8.3	522	10	US-09-765-272-120	Sequence 120, App
28	60.5	8.3	1040	10	US-09-765-272-118	Sequence 118, App
29	60	8.2	261	10	US-09-738-873-2	Sequence 2, Appli
30	60	8.2	261	10	US-09-799-777-12	Sequence 12, Appli
31	59.5	8.1	929	9	US-09-738-626-5144	Sequence 5144, Ap
32	59	8.1	117	10	US-09-864-761-37175	Sequence 37175, A
33	59	8.1	541	10	US-09-981-421-4	Sequence 4, Appli
34	59	8.1	541	12	US-10-157-447-2	Sequence 2, Appli
35	59	8.1	599	10	US-09-758-269-6	Sequence 6, Appli
36	58.5	8.0	114	10	US-09-925-300-1374	Sequence 1374, Ap
37	58.5	8.0	351	9	US-09-829-378-5	Sequence 5, Appli
38	58.5	8.0	384	9	US-09-738-626-4937	Sequence 4937, Ap
39	58.5	8.0	645	10	US-09-925-302-641	Sequence 641, App
40	58.5	8.0	855	10	US-09-828-307A-4	Sequence 4, Appli
41	58.5	8.0	865	10	US-09-881-853-2	Sequence 2, Appli
42	58.5	8.0	865	10	US-09-828-307A-2	Sequence 2, Appli
43	58	7.9	311	10	US-09-925-301-1238	Sequence 1238, Ap
44	58	7.9	354	10	US-09-841-132-576	Sequence 132, App
45	58	7.9	509	10	US-09-205-658-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-09-764-864-1181
Sequence 1181, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P1233
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1181
LENGTH: 164
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (10)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1181

Query Match 100.0%; Score 732; DB 10; Length 164;
Best Local Similarity 100.0%; Pred. No. 1.3e-76;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MASVDFKTYVDQACRAAEFVNYYTMDKRRRLSLRYLMTATLVNMGNAVSGESLSE 60

Db 25 MASVDFKTYVDQACRAAEFVNYYTMDKRRRLSLRYLMTATLVNMGNAVSGESLSE 84

Oy 61 FFEMLPSSEFOISVDDCPVDEATPSQTTVLVTCGSKFEGNKORDFNQNFILTAQAS 120

Db 85 FFEMPLSSEFQISVVDCQPVHDEATPSQTTVLVVICGSKFEGNKQDFNQNFILTAQAS 144
Qy 121 PSNTVWKIASDCFRFQDWS 140
Db 145 PSNTVWKIASDCFRFQDWS 164

RESULT 2

US-09-764-864-1180
; Sequence 1180, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1180
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1180

Query Match 79.1%; Score 579; DB 10; Length 162;
Best Local Similarity 75.7%; Pred. No. 4.9e-59;
Matches 103; Conservative 21; Mismatches 12; Indels 0; Gaps 0;

Qy 5 DFTYVDQACRAAEFVNVYTTMDKRRLLSLRYLMTATLVNGNAVSGQESLSSEFFEM 64
Db 26 DFTYVDQACRAAEFVNVYTTMDKRRLLSLRYLMTATLVNGNAVSGLDALNFFDT 85
Qy 65 LPSEFQISVVDCQPVHDEATPSQTTVLVVICGSKFEGNKQDFNQNFILTAQASPSNT 124
Db 86 LPSEFQISVVDCQPVHDEATPSQTTVLVVICGSKFEGNKQDFNQNFILTAQASPSNT 145
Qy 125 VVKIASDCFRFQDWS 140
Db 146 VVKIASDCFRFQDWS 161

RESULT 3

US-09-925-301-977
; Sequence 977, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; Prior application number: PCT/US00/05882
; Prior Filing Date: 2000-03-08
; Prior Application Number: 60/124,270
; Prior Filing Date: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 977
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (471)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (473)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-925-301-977

Query Match 14.4%; Score 105.5; DB 10; Length 477;

Best Local Similarity 28.3%; Pred. No. 0.00051;
Matches 36; Conservative 21; Mismatches 57; Indels 13; Gaps 4;
Qy 19 EFVRYVYTTMDKRRLLSLRYLMTATLVNG-----NAVSGQESLSSEFFEMPSSEF 70
Db 25 EFVRYVYTTMDKRRLLSLRYLMTATLVNG-----NAVSGQESLSSEFFEMPSSEF 81
Qy 71 QISVDCQPVHDEATPSQTTVLVVICGSKFEGNKQDFNQNFILTAQASPSNTVWKIAS 130
Db 82 TNCHTKIRHVDATLND-GVVQVVMGLLSNNQALRRFMQTVLAPGVSANKFY-VHN 139
Qy 131 DCFRQD 137
Db 140 DIFRYQD 146

RESULT 4

US-09-864-761-37945
; Sequence 37945, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; Prior Application Number: US 60/180,312
; Prior Filing Date: 2000-02-04
; Prior Application Number: US 60/207,456
; Prior Filing Date: 2000-05-26
; Prior Application Number: US 09/632,366
; Prior Filing Date: 2000-08-03
; Prior Application Number: GB 24263.6
; Prior Filing Date: 2000-10-04
; Prior Application Number: US 60/236,359
; Prior Filing Date: 2000-09-27
; Prior Application Number: PCT/US01/00666
; Prior Filing Date: 2001-01-30
; Prior Application Number: PCT/US01/00667
; Prior Filing Date: 2001-01-30
; Prior Application Number: PCT/US01/00664
; Prior Filing Date: 2001-01-30
; Prior Application Number: PCT/US01/00669
; Prior Filing Date: 2001-01-30
; Prior Application Number: PCT/US01/00665
; Prior Filing Date: 2001-01-30
; Prior Application Number: PCT/US01/00668
; Prior Filing Date: 2001-01-30
; Prior Application Number: PCT/US01/00663
; Prior Filing Date: 2001-01-30
; Prior Application Number: PCT/US01/00662
; Prior Filing Date: 2001-01-30
; Prior Application Number: PCT/US01/00661
; Prior Filing Date: 2001-01-30
; Prior Application Number: PCT/US01/00670
; Prior Filing Date: 2001-01-30
; Prior Application Number: US 60/234,687
; Prior Filing Date: 2000-09-21
; Prior Application Number: US 09/608,408
; Prior Filing Date: 2000-06-30
; Prior Application Number: US 09/774,203
; Prior Filing Date: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37945
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159
; LENGTH: 632.
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-09-205-658-159

Query Match          9.28; Score 67; DB 10; Length 632;
Best Local Similarity 22.8%; Pred. No. 20;
Matches 43; Conservative 16; Mismatches 59; Indels 72; Gaps 11;

QY      8  TTYDQACRAAEFVNVYTTMDRRRLLSRLYMGATLVWNGNAVSGQESLSEF--FEML 65
      ||: || || || || || || || || || || || || || || || || || ||
DB 122  TYLSQEC-GGHPFVTLQYTHFDQ----ARIYF-VIGLVENGDL--GESLCHFGSFDML 172

QY      66  PSSEFQISVV-----DCOPVHDEATPSTTT-----VLVTCGSKVFGNKQ----- 106
      ||: || || || || || || || || || || || || || || || || || ||
DB 173  TSFFFASEILTLGLFLHDKNKIVHRDKPDLVLIQKDHILLITDFGSAQAFGLQLSQEGF 232

QY      107  RDPNQ-----NFILTA-----QASPSNTVW 126
      || || || || || || || || || || || || || || || || || || ||
DB 233  TDANQASSRSSDSSGPPPTFRYSDEEENTARRTTFVGTALYVSPPEMLADGDVGPQTDIW 292

QY      127  KIASDCFRFQ 136
      || || || || || || || || || || || || || || || || || || ||
DB 293  GLG--CILFQ 300

RESULT 7
US-09-840-707A-3
; Sequence 3, Application US/09840707A
; Patent No. US2002007276A1
; GENERAL INFORMATION:
; APPLICANT: Fredexing, Terry M.
; APPLICANT: Ignatyev, George M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
; TITLE OF INVENTION: INFECTIONS AND OTHER DISORDERS
; FILE REFERENCE: 24891-301C
; CURRENT APPLICATION NUMBER: US/09/840,707A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/562,979
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/198,210
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Interleukin-1 receptor, Type I precursor
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: P14778/GenBank
; DATABASE ENTRY DATE: 2001-08-20
; US-09-840-707A-3

Query Match          9.1%; Score 66.5; DB 10; Length 569;
Best Local Similarity 22.5%; Pred. No. 20;

```

[illegible]

RESULT 8
US-09-731-175-2
Sequence 2, Application US/097311175
Patent No. US20020098169A1
GENERAL INFORMATION:
APPLICANT: Glorioso, Joseph C.
Evans, Christopher H.
Robbins, Paul D.
TITLE OF INVENTION: Gene Transfer for Studying and Treating
a Connective Tissue of a Mammalian Host
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/731,175
FILING DATE: 05-Dec-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER:	08/924,777
FILING DATE:	2000-01-31
APPLICATION NUMBER:	US 07/963,928
FILING DATE:	20-OCT-1992
APPLICATION NUMBER:	US 08/027,750
FILING DATE:	08-MAR-1993
APPLICATION NUMBER:	US 08/183,563
FILING DATE:	18-JAN-1994
APPLICATION NUMBER:	US 08/381,603
FILING DATE:	27-JAN-1995
APPLICATION NUMBER:	US 08/567,710
FILING DATE:	05-DEC-1995
APPLICATION NUMBER:	US 08/685,212
FILING DATE:	23-JUL-1996

- ATTORNEY/AGENT INFORMATION:
 NAME: Bastian, Kevin L.
 REGISTRATION NUMBER: 34,774
 REFERENCE/DOCKET NUMBER: 018484-002280US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 569 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-731-175-2

Query Match 9.1%; Score 66.5; DB 10; Length 569;
Best Local Similarity 22.5%; Pred. No. 20;
Matches 25; Conservative 20; Mismatches 53; Indels 13; Gaps 5;

```

QY 18 EEEVNVYTTMDKRRLLSLRYMGTTATLVYNGNAVSGQESLSEFFEMLPSSSEFQISVVD 77
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 434 EDIVVINENKSKRLLIILRETSGFSLWG--SSEQIA-MYNALVQDGIKVLLLEL 490
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 78 QPVHD-EATPSQTTVLVVICGSKVFGNKQKORDNQNFILTAQASPSNTVWK 127
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 EKIQDYEKMPESIKFKQKHGAIRWGS---DFTQ-----GPOSAKTRPWK 532
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 9
US-09-205-658-160
; Sequence 160, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNO
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLER
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/957,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 160
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-205-658-160

```

	Query Match	9.18;	Score 66.5;	DB 10;	Length 636;	
	Best Local Similarity	28.1%;	Pred. No. 23;	Mismatches 53;	Indels 25;	Gaps 8;
	Matches 36;	Conservative 14;				
Qy	8 TYVDQAARAAEEFNVVYTTMDKKRRLLSRLYMGTATLVWNGNAVSGOESLSEF--FEML 65 	:	: :	::	::	.: .
Dd	122 TYLSEOEC-GGHPFTQLATHFDQ---ARIYF-VIGLVENGDL---GESLSCHTGSFML 172 	:	: :	::	::	.: .
Qy	66 PSSSFQISIW-----DCQPVPHEDEATPSOTT-----VLVVICGSVFKEGNKQRDFNON 112 :	:	: :	::	::	.: .
Dd	173.TSKFFASILLGLQFLHDNKIVHRDMKDKNVLIQDKGHILITDSCSAQAFGLQLS-QEG 231 :	:	: :	::	::	.: .
Qy	113 FILTAQAS 120 	:	:	:	:	:
Dd	232 ETDANOAS 239 	:	:	:	:	:

```

RESULT 10
US-09-303-510-6
; Sequence 6, Application US/09303510A
; Patent No. US20020028209A1
; GENERAL INFORMATION:
; APPLICANT: Collisson, Ellen W.
; APPLICANT: Hash, Stephen M.
; APPLICANT: Choi, InSo
; TITLE OF INVENTION: Feline CD80, Feline CD86, Feline CD28, and Feline
; CD137 Nucleic Acid and Polypeptides
; FILE REFERENCE: 54954
; CURRENT APPLICATION NUMBER: US/09/303,510A
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,869
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: Patentin Ver. 2.1.
; SEQ ID NO 6
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Feline

```

US-09-303-510-6

Query Match 8.9%; Score 65; DB 10; Length 329;

Best Local Similarity 27.3%; Pred. No. 14;

Matches 30; Conservative 14; Mismatches 42; Indels 24; Gaps 5;

QY 18 EEFVNVYTTMDKRRLLSRILYMGATATLVWNGNAVSGOESLSEF-----FEMLPSEF 70

Db 178 ENSTTKYDVMKKSQNNVTLYNYSISLPF---SVPEAHNVSVFCALKLETLEMLSLPF 234

QY 71 QISVVDQCPVHDEATPSQ-----TTVLVVICGSKFEGNKQRFQ 111

Db 235 NI---DAQP---KDKDPEGHFLWIAAFLVMEVFCGMVSFKTLRKRKKQ 279

RESULT 11

US-09-303-040-6

Sequence 6, Application US/09303040

Patent No. US20020051792A1

GENERAL INFORMATION:

APPLICANT: Winslow, Barbara J.

APPLICANT: Cochran, Mark D.

TITLE OF INVENTION: Recombinant Virus Expressing Foreign DNA Encoding

TITLE OF INVENTION: Feline CD80, Feline CD86, Feline CD28, Feline CTLA-4 or

TITLE OF INVENTION: Feline Interferon-gamma And Uses Thereof

FILE REFERENCE: 54957-B

CURRENT APPLICATION NUMBER: US/09/303,040

CURRENT FILING DATE: 1999-04-30

EARLIER APPLICATION NUMBER: 60/083,870

EARLIER FILING DATE: 1998-05-01

NUMBER OF SEQ ID NOS: 82

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 6

LENGTH: 329

TYPE: PRT

ORGANISM: feline CD86

US-09-303-040-6

Query Match

Best Local Similarity 8.9%; Score 65; DB 10; Length 329;

Matches 30; Conservative 14; Mismatches 42; Indels 24; Gaps 5;

QY 18 EEFVNVYTTMDKRRLLSRILYMGATATLVWNGNAVSGOESLSEF-----FEMLPSEF 70

Db 178 ENSTTKYDVMKKSQNNVTLYNYSISLPF---SVPEAHNVSVFCALKLETLEMLSLPF 234

QY 71 QISVVDQCPVHDEATPSQ-----TTVLVVICGSKFEGNKQRFQ 111

Db 235 NI---DAQP---KDKDPEGHFLWIAAFLVMEVFCGMVSFKTLRKRKKQ 279

RESULT 12

US-09-758-269-12

Sequence 12, Application US/09758269

Patent No. US20020104120A1

GENERAL INFORMATION:

APPLICANT: IUCHI, SATOSHI

APPLICANT: KOBAYASHI, MASATOMO

APPLICANT: SHINOZAKI, KAZUO

TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

TITLE OF INVENTION: CLEAVAGE ENZYME GENE

FILE REFERENCE: 3914-3

CURRENT APPLICATION NUMBER: US/09/758,269

CURRENT FILING DATE: 2001-01-12

PRIOR APPLICATION NUMBER: JP 2001-003476

PRIOR FILING DATE: 2001-01-11

PRIOR APPLICATION NUMBER: JP 2000-010056

PRIOR FILING DATE: 2000-01-13

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 12

LENGTH: 612

TYPE: PRT

ORGANISM: Vigna unguiculata

US-09-758-269-12

Query Match

Best Local Similarity 8.9%; Score 65; DB 10; Length 612;

Matches 28; Conservative 14; Mismatches 39; Indels 20; Gaps 6;

QY 55 QESLSEFFEMLP-----SSEFOISVVDQCPVHDEA-----TPSQTTVL-----VVICGS- 98

Db 332 QKPYLKVFERSPDGVKSPDVEIPLKEPTMHDFAITENFVVVPDQVVFKLTENITGSP 391

QY 99 VKFEGNKQRFQNFILTAQASPSNTV-WKIASDCFRFQDW 138

Db 392 VVYDKNKTREG---ILHNAKDANAMRWIDAPDCFCFHLW 429

RESULT 13

US-10-155-613-2

Sequence 2, Application US/10155613

Publication No. US20020192787A1

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

APPLICANT: Corley, Neil C.

APPLICANT: Guegler, Karl J.

APPLICANT: Yue, Henry

TITLE OF INVENTION: HUMAN NUCLEIC ACID METHYLASES

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/155,613

FILING DATE: 22-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION NUMBER: US/09/375,205

FILING DATE: <Unknown>

APPLICATION NUMBER: 09/082,310

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CERRONE, MICHAEL C.

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PF-0520 US

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 396 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: THPLAZS08

CLONE: 2757184

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-155-613-2

Query Match

Best Local Similarity 8.7%; Score 64; DB 9; Length 396;

Matches 35; Conservative 17; Mismatches 62; Indels 50; Gaps 5;

QY 2 ASVDFKTYVDQACRAAEFNVVYTTMDKRRLLSRILYMG-----TATLVWNGNAVSGQES 57

```

Db 67 ASDFKRYVDR-RLAETLAQIYLGKPSRPPHLLLECNPGGILTOALLEAGAKVVALES 123
QY 58 LSEFFEMLPs-----SEFQISVWDC-----OPVHDEATPSQT 89
Db 126 DKTFIPHLESGLGNLDGKLRVHCDFKLDPRGGVTKPPAMSSRGLFKNLGIEAVPMTA 185
QY 90 TVLVVIGSVKFEKNKQDFNQNFILTAQASPSNTVWKIASDCF 133
Db 186 DIPLKVVGMFPSPRGEK-----ALWKLAYDLY 212

RESULT 14
US-09-925-301-1057
; Sequence 1057, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1057
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (343)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1057

Query Match 8.7%; Score 64; DB 10; Length 407;
Best Local Similarity 21.3%; Pred. No. 24;
Matches 35; Conservative 17; Mismatches 62; Indels 50; Gaps 5;

QY 2 ASVDKRYVDOACRAAEFNVYTTMDKRRRLSLRYMG-----TATLVNNGNAVSGQES 57
Db 78 ASDFKRYVDR-RLAETLAQIYLGKPSRPPHLLLECNPGGILTOALLEAGAKVVALES 136
QY 58 LSEFFEMLPs-----SEFQISVWDC-----OPVHDEATPSQT 89
Db 137 DKTFIPHLESGLGNLDGKLRVHCDFKLDPRGGVTKPPAMSSRGLFKNLGIEAVPMTA 196
QY 90 TVLVVIGSVKFEKNKQDFNQNFILTAQASPSNTVWKIASDCF 133
Db 197 DIPLKVVGMFPSPRGEK-----ALWKLAYDLY 223

RESULT 15
US-10-058-566-8
; Sequence 8, Application US/10058566
; Publication No. US20020183274A1
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, MARC
; APPLICANT: ROX, TIM
; APPLICANT: HUFFMAN, GARY
; APPLICANT: TRIMMELL, MARY
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND
; TITLE OF INVENTION: METHOD OF USING SAME
; FILE REFERENCE: PHI 1147
; CURRENT APPLICATION NUMBER: US/10/058,566
; CURRENT FILING DATE: 2002-01-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 8
; LENGTH: 232
```

```

; TYPE: PRT
; ORGANISM: Sorghum sp.
US-10-058-566-8

Query Match 8.7%; Score 63.5; DB 9; Length 232;
Best Local Similarity 20.2%; Pred. No. 13;
Matches 20; Conservative 16; Mismatches 40; Indels 23; Gaps 3;

QY 63 EMLPSSEFQISVVDQCPVHDEATP-----SQTTLVYVIGSVK-FEGNKQDFNQ 112
Db 55 DLLEGSFDNAVMDGCVFHTASPVLAKESSSSKEETLCPAVNGTLNVLRSCKKNPELKR 114
QY 113 FILTAQASP-----SNTVWKIASDCFRFQDW 138
Db 115 VLTSSSAVRIRDDQDPNISLDETTWSSVPLCEKMLW 153

Search completed: March 4, 2003, 15:07:07
Job time : 4.35122 secs
```

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:50:41 ; Search time 44.3388 Seconds
(without alignments)
1559.740 Million cell updates/sec

Title: US-09-763-902b-3

Perfect score: 2620

Sequence: 1 MAAREPPPLGGKPTDFED.....TQQLIKYWEAFLPEAKATA 519

Scoring table: BLOSUM62

Gap 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2620	100.0	519	21	AA82319 Human protein tran
2	2620	100.0	519	22	AA82319 Human protein sequ
3	2616	99.8	519	22	AA82319 Human protein sequ
4	2611	99.7	519	22	AA82319 Human protein sequ
5	2033	77.6	451	19	AAW7095 Sorting nexin 2.
6	1554.5	59.3	522	19	AAW7095 Sorting nexin 1.
7	1062	40.5	458	22	AA859331 Drosophila melanog
8	397.5	15.2	554	21	AA859331 Arabidopsis thalia
9	397.5	15.2	571	21	AA859331 Arabidopsis thalia
10	397.5	15.2	572	21	AA859331 Arabidopsis thalia

11	368	14.0	586	21	AA841594 Arabidopsis thalia
12	368	14.0	587	21	AA841594 Arabidopsis thalia
13	367	14.0	566	21	AA841595 Arabidopsis thalia
14	345.5	13.2	314	21	AA827811 Arabidopsis thalia
15	345.5	13.2	327	21	AA827810 Arabidopsis thalia
16	345.5	13.2	342	21	AA827809 Arabidopsis thalia
17	337	12.5	68	22	ABB43886 Protein #807 enco
18	327	12.5	68	22	ABB43886 Protein #807 enco
19	327	12.5	68	22	AA64876 Human brain expres
20	327	12.5	68	22	AA64876 Human bone marrow
21	327	12.5	68	22	AA64876 Peptide #7972 enco
22	327	12.5	68	22	AA64876 Peptide #11845 enco
23	327	12.5	68	23	ABC46647 Human peptide enco
24	327	12.5	455	22	ABB11613 Human sorting nexi
25	324.5	12.4	443	22	AAO13502 Human polypeptide
26	324.5	12.4	443	22	AAO13502 Human polypeptide
27	324	12.4	447	22	AA841594 Human colon cancer
28	308.5	11.8	404	22	AA841594 Human protein sequ
29	306	11.7	61	22	AAU17428 Novel signal trans
30	293.5	11.2	404	22	AAU17428 Novel human secret
31	281	10.7	403	23	ABG66885 Human novel polype
32	281	10.7	403	23	ABG66885 Human novel polype
33	257.5	9.8	428	22	ABB63771 Drosophila melanog
34	248	9.5	406	21	AA841594 Human TRAF four as
35	248	9.5	414	21	AA841594 Lung cancer associ
36	248	9.5	414	21	AA841594 Amino acid sequenc
37	248	9.5	414	23	ABP41711 Human ovarian anti
38	243.5	9.3	460	21	AA841594 Human ORFX ORF2921
39	237.5	9.1	211	21	AA841594 Human ORFX ORF2956
40	221	8.4	450	22	AA841594 Human protein sequ
41	221	8.4	450	22	AA841594 SNEXXN protein #2
42	205.5	7.8	463	21	AA841594 Human ORFX ORF2906
43	200	7.6	595	21	AA841594 Human ORFX ORF3067
44	190	7.3	565	22	ABP51340 Drosophila melanog
45	187.5	7.2	543	23	ABP51340 Human MDDT S80 ID

ALIGNMENTS

RESULT 1
AA82319
ID: AA82319 standard; Protein; 519 AA.
XX AA82319;
XX 19-JUN-2000 (first entry)
DE Human protein transport molecule (PTAM) SEQ ID NO:3.
XX
KW Human; protein transport molecule; PTAM; diagnosis; cytostatic;
KW antithratic; antithratic; immunosuppressant; antithratic;
KW antiallergic; antidiabetic; antilipemic; antirheumatic; osteoporotic;
KW dermatological; antianemic; antipsoriatic; hepatotropic; antigout;
KW antinflammatory; antihiv; protein transport regulator; cancer;
KW immune disorder; cell proliferative disorder; secretory disorder;
KW articular; allergy; abnormal vesicle trafficking; asthma;
KW autoimmune haemolytic anaemia.

OS Homo sapiens.

PN WO200012703-A2.

XX 09-MAR-2000.

XX 26-AUG-1999; 99WO-US19616.

XX 27-AUG-1998; 98US-0098206.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;

PI Gorgone GA, Baughn MR, Patterson C;

XX WPI; 2000-256642/22.
 DR N-PSDB; AAA08037.
 XX
 PT New human protein transport-associated polypeptide and polynucleotide
 PT useful for diagnosis, prevention and treatment of cell proliferative
 PT and secretory disorders such as leukemia, cystic fibrosis
 XX
 PS Claim 1; Page 62-63; 75pp; English.
 XX
 CC AAA08035 to AAA08042 encode the human protein transport-associated
 CC molecules (PTAMs) given in AAY82317 to AAY82324. The PTAMs have
 CC cytotactic, antiarthritic, antiasthmatic, immunosuppressant,
 CC antitumor, antiallergic, antidiabetic, antilipemic,
 CC antirheumatic, osteopathic, dermatological, antianemic, antipsoriatic,
 CC hepatotropic, antigout, antiinflammatory and antihiv activities, and
 CC regulate protein transport. PTAM proteins and antagonists are useful for
 CC preventing or treating a disorder associated with decreased or increased
 CC expression or activity of PTAM. PTAM polynucleotides are useful for
 CC diagnosing conditions associated with PTAM, comprising detecting PTAM by
 CC forming a hybridisation complex, preferably after PCR amplifying the
 CC biological sample. Diseases prevented, treated or diagnosed include cell
 CC proliferative disorders such as cancers, immune disorders, secretory
 CC disorders and other conditions associated with abnormal vesicle
 CC trafficking, such as allergies, asthma, urticaria and autoimmune
 CC haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as
 CC a targeting or delivery mechanism for bringing pharmaceutical agents
 CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related
 CC disorders. PTAM, its catalytic or immunogenic fragments are useful for
 CC drug screening using libraries of compounds. PTAM polynucleotides are
 CC useful for generating hybridisation probes useful in mapping the
 CC naturally occurring genomic sequences.
 XX Sequence 519 AA;

Query Match 100.0%; Score 2620; DB 21; Length 519;
 Best Local Similarity 100.0%; Pred. No. 4.7e-198;
 Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAREPPPLGDKPTDFEDLGEDLFTSTVSTLESSPSPSPASLPADISANSNGPK 60
 DB 1 MAAREPPPLGDKPTDFEDLGEDLFTSTVSTLESSPSPSPASLPADISANSNGPK 60
 QY 61 PTEVLDDREDFATEEVSLSDEPERPILSSPSPAVTPTPTLLAPRIESKMSA 120
 DB 61 PTEVLDDREDFATEEVSLSDEPERPILSSPSPAVTPTPTLLAPRIESKMSA 120
 QY 121 PVIFDRSREIEEANGDIFDIEIGVSDPEKVGDNAYMAYRVTTKTSLSMFSKSEFSV 180
 DB 121 PVIFDRSREIEEANGDIFDIEIGVSDPEKVGDNAYMAYRVTTKTSLSMFSKSEFSV 180
 QY 181 KRRFSDFLGHSKLASKYLHVGYIVPPAPEKSIIVGMYKVKGEKDESSSTFEVKKRAALE 240
 DB 181 KRRFSDFLGHSKLASKYLHVGYIVPPAPEKSIIVGMYKVKGEKDESSSTFEVKKRAALE 240
 QY 241 RYLQRTVKIPTLQDPLDRLQFLESSELPRVNTQALSGAGILMNVNKAADAVNKTIKN 300
 DB 241 RYLQRTVKIPTLQDPLDRLQFLESSELPRVNTQALSGAGILMNVNKAADAVNKTIKN 300
 QY 301 ESDAWEFEKQFENDQQLKHLVSEALVCHRKELSAANTAFKSAAMLGNSEDTAL 360
 DB 301 ESDAWEFEKQFENDQQLKHLVSEALVCHRKELSAANTAFKSAAMLGNSEDTAL 360
 QY 361 SRALSQALAEVEEKIDQLHQEQAFADYMFSELLSDYIRLIAAIVKGVFDRHMKCWQKWEA 420
 DB 361 SRALSQALAEVEEKIDQLHQEQAFADYMFSELLSDYIRLIAAIVKGVFDRHMKCWQKWEA 420
 QY 421 QITLLKREAEAKMYNPKDQQAQKNEIREWEAKVQOGERDQFQISKTIRKEVGRFEK 480
 DB 421 QITLLKREAEAKMYNPKDQQAQKNEIREWEAKVQOGERDQFQISKTIRKEVGRFEK 480
 QY 481 ERVKDFKTVIIKYLESLVQTQQOLIKYWEAFLPEAKAIA 519
 DB 481 ERVKDFKTVIIKYLESLVQTQQOLIKYWEAFLPEAKAIA 519

DB 481 ERVKDFKTVIIKYLESLVQTQQOLIKYWEAFLPEAKAIA 519
 RESULT 2
 AAB95444
 ID AAB95444 standard; Protein; 519 AA.
 XX AAB95444;
 AC AAB95444;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17891.
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX EPI074617-A2.
 PN 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 PF 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 17891; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 519 AA;
 SQ
 Query Match 100.0%; Score 2620; DB 22; Length 519;
 Best Local Similarity 100.0%; Pred. No. 4.7e-198;
 Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAREPPPLGDKPTDFEDLGEDLFTSTVSTLESSPSPSPASLPADISANSNGPK 60

```
Db 1 MAAREPPPLGDKPTDFEDLEDGEDLFTSTVSTLESSPSPASPASLPADISANSNGPK 60
Qy 61 PTEVLDLDDREDFLAEATEEVSLSLSPEREPILSSPSPAVTPTPTTLIAPRIEKSMSA 120
Db 61 PTEVLDLDDREDFLAEATEEVSLSLSPEREPILSSPSPAVTPTPTTLIAPRIEKSMSA 120
Qy 121 PVIFDRSREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSV 180
Db 121 PVIFDRSREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSV 180
Qy 181 KRRFSDFLGLHSLKASKYLHVGVIYPPPAPEKSIVGMTKVKGKEDSSSTFEVKKRAALE 240
Db 181 KRRFSDFLGLHSLKASKYLHVGVIYPPPAPEKSIVGMTKVKGKEDSSSTFEVKKRAALE 240
Qy 241 RYLQRTVKHPTLLQDPDLRQFLLESSLPRAVNTQALSGAGILRMVKNKAADAVNKMTKNN 300
Db 241 RYLQRTVKHPTLLQDPDLRQFLLESSLPRAVNTQALSGAGILRMVKNKAADAVNKMTKNN 300
Qy 301 ESDAWFEKQOQFENLDQOLRLHVSVEALVCHRKELSANTAAFAKSAAMLGNSDHTAL 360
Db 301 ESDAWFEKQOQFENLDQOLRLHVSVEALVCHRKELSANTAAFAKSAAMLGNSDHTAL 360
Qy 361 SRALSOLAEVEEKIDOLHGOEAFADYMFSELLSDYIRLIAAVKGVDFHRMKCQWKEDA 420
Db 361 SRALSOLAEVEEKIDOLHGOEAFADYMFSELLSDYIRLIAAVKGVDFHRMKCQWKEDA 420
Qy 421 QITLLKKREAEAKMMVANKPKDQIQAKNEIREWEAKVQOGERDFEQISKTIRKEVGRFEK 480
Db 421 QITLLKKREAEAKMMVANKPKDQIQAKNEIREWEAKVQOGERDFEQISKTIRKEVGRFEK 480
Qy 481 ERVKDFKTVIIKYLESLVOTQOOLIKYWEAFLEPEAKAIA 519
Db 481 ERVKDFKTVIIKYLESLVOTQOOLIKYWEAFLEPEAKAIA 519
RESULT 3
AAB93676
ID AAB93676 standard; Protein; 519 AA.
XX AAB93676;
AC AAB93676;
DT 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:13214.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs
```

```
XX Claim 8; SEQ ID 13214; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX Sequence 519 AA;
SQ
Query Match 99.8%; Score 2616; DB 22; Length 519;
Best Local Similarity 99.8%; Pred. No. 9.7e-198;
Matches 518; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAAREPPPLGDKPTDFEDLEDGEDLFTSTVSTLESSPSPASPASLPADISANSNGPK 60
Db 1 MAAREPPPLGDKPTDFEDLEDGEDLFTSTVSTLESSPSPASPASLPADISANSNGPK 60
Qy 61 PTEVLDLDDREDFLAEATEEVSLSLSPEREPILSSPSPAVTPTPTTLIAPRIEKSMSA 120
Db 61 PTEVLDLDDREDFLAEATEEVSLSLSPEREPILSSPSPAVTPTPTTLIAPRIEKSMSA 120
Qy 121 PVIFDRSREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSV 180
Db 121 PVIFDRSREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSV 180
Qy 181 KRRFSDFLGLHSLKASKYLHVGVIYPPPAPEKSIVGMTKVKGKEDSSSTFEVKKRAALE 240
Db 181 KRRFSDFLGLHSLKASKYLHVGVIYPPPAPEKSIVGMTKVKGKEDSSSTFEVKKRAALE 240
Qy 241 RYLQRTVKHPTLLQDPDLRQFLLESSLPRAVNTQALSGAGILRMVKNKAADAVNKMTKNN 300
Db 241 RYLQRTVKHPTLLQDPDLRQFLLESSLPRAVNTQALSGAGILRMVKNKAADAVNKMTKNN 300
Qy 301 ESDAWFEKQOQFENLDQOLRLHVSVEALVCHRKELSANTAAFAKSAAMLGNSDHTAL 360
Db 301 ESDAWFEKQOQFENLDQOLRLHVSVEALVCHRKELSANTAAFAKSAAMLGNSDHTAL 360
Qy 361 SRALSOLAEVEEKIDOLHGOEAFADYMFSELLSDYIRLIAAVKGVDFHRMKCQWKEDA 420
Db 361 SRALSOLAEVEEKIDOLHGOEAFADYMFSELLSDYIRLIAAVKGVDFHRMKCQWKEDA 420
Qy 421 QITLLKKREAEAKMMVANKPKDQIQAKNEIREWEAKVQOGERDFEQISKTIRKEVGRFEK 480
Db 421 QITLLKKREAEAKMMVANKPKDQIQAKNEIREWEAKVQOGERDFEQISKTIRKEVGRFEK 480
Qy 481 ERVKDFKTVIIKYLESLVOTQOOLIKYWEAFLEPEAKAIA 519
Db 481 ERVKDFKTVIIKYLESLVOTQOOLIKYWEAFLEPEAKAIA 519
RESULT 4
AAB94707
ID AAB94707 standard; Protein; 519 AA.
XX
```


AC AAB94707;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:15700.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 XX EP1074617-A2.
 PN
 XX
 PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR
 PR 27-AUG-1999; 99JP-0300253.
 PR
 PR 11-JAN-2000; 2000JP-0118776.
 PR
 PR 02-MAY-2000; 2000JP-0183767.
 PR
 PR 09-JUN-2000; 2000JP-0241899.
 PR
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 PI
 XX WPI; 2001-318749/34.
 DR
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 15700; 2537pp + CD ROM; English.
 PS
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 519 AA;
 SQ
 Query Match 99.7%; Score 2611; DB 22; Length 519;
 Best Local Similarity 99.6%; Pred. No. 2.4e-197;
 Matches 517; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 MAAREPPPLGDKPTDFDLEDGEDLFTSTVSTLESSPSPPEASLPADIESANSNGPK 60
 DB 1 MAAREPPPLGDKPTDFDLEDGEDLFTSTVSTLESSPSPPEASLPADIESANSNGPK 60
 QY 61 PTEVLDDREDLFARATEVSLDSPEREPILSSPSPAVTPTTTLAPRIEKSMSA 120
 DB 61 PTEVLDDREDLFARATEVSLDSPEREPILSSPSPAVTPTTTLAPRIEKSMSA 120

QY 121 PVIFDRREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTTKTSLSMFSKSEFSV 180
 DB 121 PVIFDRREEIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTTTKTSLSMFSKSEFSV 180
 QY 181 KRRFSDFLGLHSLKSLKYLHVGVIYVPPAPEKSIIVGKTKVVKGKEDSSSTFEFVEKRRAALE 240
 DB 181 KRRFSDFLGLHSLKSLKYLHVGVIYVPPAPEKSIIVGKTKVVKGKEDSSSTFEFVEKRRAALE 240
 QY 241 RYLQRTVKHPTLLQDPDLRQFLFLESSELPRAVNTQALSGAGILRMVKNKAADAVNKMVTIKMN 300
 DB 241 RYLQRTVKHPTLLQDPDLRQFLFLESSELPRAVNTQALSGAGILRMVKNKAADAVNKMVTIKMN 300
 QY 301 ESDAWFEKQOQFENLDQOLRLKLVSVLCHRELSANTAAFAKSAAMLGNSEDHTAL 360
 DB 301 ESDAWFEKQOQFENLDQOLRLKLVSVLCHRELSANTAAFAKSAAMLGNSEDHTAL 360
 QY 361 SRALSQAEVEEKIDQLHQEAFADFYMFSELLSDYIRLIAAVKGVDFHRMCKWKWEDA 420
 DB 361 SRALSQAEVEEKIDQLHQEAFADFYMFSELLSDYIRLIAAVKGVDFHRMCKWKWEDA 420
 QY 421 QITLLKREAEAKMVANKPKDKIQQAKNEIREWEAKVQOGERDFEQISKTIRKEVGRPEK 480
 DB 421 QITLLKREAEAKMVANKPKDKIQQAKNEIREWEAKVQOGERDFEQISKTIRKEVGRPEK 480
 QY 481 ERVKDFKTVIKYLSLVQTQOOLIKYWEAFLEPAKAI 519
 DB 481 ERVKDFKTVIKYLSLVQTQOOLIKYWEAFLEPAKAI 519
 RESULT 5
 AAW77095
 ID AAW77095 standard; Protein: 451 AA.
 XX
 AC AAW77095;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Sorting nexin 2.
 XX
 KW Human; sorting nexin; intracellular domain; cell surface receptor;
 KW translocation; lysosome; down-regulation; cancer.
 XX
 OS Homo sapiens.
 XX
 PN US3804412-A.
 XX
 PD 08-SEP-1998.
 XX
 PF 01-APR-1996; 96US-0625322.
 XX
 PR 01-APR-1996; 96US-0625322.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Cadena DL, Gill GN, Kurten RC;
 XX
 DR WPI; 1998-505652/43.
 DR N-PSDB; AAV48249.
 XX
 XX Nucleic acids encoding sorting nexin proteins - used for recombinant
 PT production and regulating expression of the protein, especially in
 PT treating unregulated cell growth e.g. cancer.
 XX
 PS Claim 5; Column 29-34; 25pp; English.
 XX
 CC Sorting nexins (SNX) have the ability to bind an intracellular domain of
 CC an activated cell surface receptor expressed on a cell and direct
 CC translocation of the receptor to a lysosome in the cell. The nucleic
 CC acids encoding SNX are useful for the down-regulation of cell surface
 CC receptors for degradation when they are no longer required to carry out a
 CC signal. The nucleic acids can be used to express the protein in a cell
 CC where expression is desired e.g. where cancer is caused by up-regulation
 CC of receptors. Antisense constructs can be used to inhibit native

CC expression of the protein.

XX Sequence 451 AA;

Query Match 77.6%; Score 2033; DB 19; Length 451;

Best Local Similarity 90.0%; Pred. No. 7.7e-152;

Matches 416; Conservative 7; Mismatches 17; Indels 22; Gaps 4;

QY 68 DREDLFAATEEVSIDS-----PEREILSPSPAVTPVPTTLI---APRIEKS 117

DB 2 DREDLFAATEEVSIDLGKLNLPQLQSH-----LSLLHSAPRIEKS 50

QY 118 MSAPVIFORSREIEEANGDIFDIIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSE 177

DB 51 MSAPVIFORSREIEEANGDIFDIIEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSE 110

QY 178 FSVKRFSDFLGHSKLSKYLHVGIYVPPAPEKSIYVGMTKVKVGKEDSSSTFEVEKRA 237

DB 111 FSVK-RETDFLGLHTLPTTYLHVIFVATSRKSIYVGMTKVKVGKEDSSSTFEVEKRA 169

QY 238 ALERYLQRTVKHPTLLQDPLDQRFLESSELPRAVNTQALSGAGILRWVNKAADAVNKMTI 297

DB 170 ALERYLQRTVKHPTLLQDPLDQRFLESSELPRAVNTQALSGAGILRWVNKAADAVNKMTI 229

QY 298 KKNESDANFEERQOQFENLDQQLRLKLVSVREALVCHRKELSANTAFKSAAMLGNSDEH 357

DB 230 KKNESDANFEERQOQFENLDQQLRLKLVSVREALVCHRKELSANTAFKSAAMLGNSDEH 289

QY 358 TALSRLSOLAEVEEKIDQLHQAQFADYMFSELSLSDYIRLIAAVKGVDFDHMKCQKW 417

DB 290 TALSRLSOLAEVEEKIDQLHQAQFADYMFSELSLSDYIRLIAAVKGVDFDHMKCQKW 349

QY 418 EDAQITLLKRAEAKMVAANKPDKIQQAQNEIREWEAKVQOGERDFEIOISKIRKEVR 477

DB 350 EDAQITLLKRAEAKMVAANKPDKIQQAQNEIREWEAKVQOGERDFEIOISKIRKEVR 409

QY 478 FEKERVKDFKTVIIRKYLESLVTOQQLIKYWEAFLEAKAIA 519

DB 410 FEKERVKDFKTVIIRKYLESLVTOQQLIKYWEAFLEAKAIA 451

RESULT 6

ID AA777094

XX AA777094 standard; Protein; 522 AA.

AC AA777094;

XX 09-NOV-1998 (first entry)

XX Sorting nexin 1.

XX Human; sorting nexin; intracellular domain; cell surface receptor;

KW translocation; lysosome; down-regulation; cancer.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 115

FT /note= "Xaa not specified, encoded by Tyr"

XX US5804412-A.

XX 08-SEP-1998.

XX 01-APR-1996; 9605-0625322.

XX 01-APR-1996; 9605-0625322.

XX (REGC) UNIV CALIFORNIA.

XX Cadena DL, Gill GN, Kurten RC;

XX WPI; 1998-505652/43.

DR

DR N-PSDB; AAV48248.

XX Nucleic acids encoding sorting nexin proteins - used for recombinant production and regulating expression of the protein, especially in treating unregulated cell growth e.g. cancer

XX Claim 3; Column 23-28; 25pp; English.

XX Sorting nexins (SNX) have the ability to bind an intracellular domain of an activated cell surface receptor expressed on a cell and direct translocation of the receptor to a lysosome in the cell. The nucleic acids encoding SNX are useful for the down-regulation of cell surface receptors for degradation when they are no longer required to carry out a signal. The nucleic acids can be used to express the protein in a cell where expression is desired e.g. where cancer is caused by up-regulation of receptors. Antisense constructs can be used to inhibit native expression of the protein.

XX Sequence 522 AA;

Query Match 59.3%; Score 1554.5; DB 19; Length 522;

Best Local Similarity 59.6%; Pred. No. 5.2e-114;

Matches 319; Conservative 78; Mismatches 99; Indels 39; Gaps 8;

QY 2 AAREPPPL-----GDGKPTDFEDLEDGEDFTTSTVLTLESSSPSE--PASLP 48

DB 10 ASERLPPPPGLEPESEGAAGGSEPEAGSDTEGEDIPTG--AAVVKHQSPKITTSLLP 67

QY 49 AEDISANSNGPPTTEVLDDRE--DLFAATEEVSLSDE--REPILSSESPAVTPVT 104

DB 68 I-----NNGSKENGIEHQDQEPQDLFADATVELSLDSTQNKKVLAK----- 111

QY 105 PTLTIAPRTESKSMAPVIFDRSREIEEANGDIFDIIEIGVSDPEKVGDMNAYMAYRV 164

DB 112 -TLIXLPQEAINSSKP---QPTYEELEEEEDQDPLTVGTIDPEKIDGGMNAYVAVK 167

QY 165 TTKTSLSMFSKSEFSYKRRFSDFLGLHSLKASKYLHVGIYVPPAPEKSIYVGMTKVKVGKE 224

DB 168 TTQTSPLPFRSQFAVKKRRFSDFLGLYKLSKHSQNGFIVPPPEKSLIGMTKVKVGKE 227

QY 225 DSSSTFEVEKRAALERYLQRTVKHPTLLQDPLDQRFLESSELPRAVNTQALSGAGILRW 284

DB 228 DSSAEFLKRAALERYLQRTVNHPTLQDPLDQRFLEKEELPRAVGTQTLGAGLLKM 287

QY 285 VNKAADAVNKMTIKNESDAWFEKQOQFENLDQQLKLVSVREALVCHRKELSANTAAP 344

DB 288 FNKATDAVSKMTIKNESDIWFEKQOQFENLDQQLKLVSVREALVCHRKELSANTAAP 347

QY 345 AKSAAMLGNSDEHTALSRLSOLAEVEEKIDQLHQAQFADYMFSELSLSDYIRLIAAVK 404

DB 348 AKSLAMLGSSDNTALSRLSOLAEVEEKIEQLHQAQFADYMFSELSLSDYIRLIAAVK 407

QY 405 GVFDRMKCQKWEDAQITLLKRAEAKMVAANKPDKIQQAQNEIREWEAKVQOGERDF 464

DB 408 AAFDQRMKTQWQDAQATLQKKRAEARLLWANKPDKLQQAQNEIREWEAKVQOGERDF 467

QY 465 EQISKTRKEVGFVEKERVKDFKTVIIRKYLESLVTOQQLIKYWEAFLEAKAIA 519

DB 468 ERISTYVRKEVIRFEKSKDFKNHVIKYLETLLYSQQLAKYWEAFLEAKAIA 522

RESULT 7

ID ABB59331 standard; Protein; 458 AA.

XX ABB59331;

XX AC ABB59331;

XX DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 4785.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

KW

Db 61 ESRSRPNKNGRRVSYLPPS-----YADVIFSPDDISEINGSEDGHS 106
Qy 138 DIFD-----IEGVSDPEKVGDMNA-----YMAVRVTKTSLSMFSKSE 177
Db 107 QSDSLSRSPSSLSDDYIKITVSNPQKEQATNSMIPGGSTIYITITRNLSLDYGGSE 166
Qy 178 FSVKRRFSDGLHSLKSLKSLYHGVIVPAPPEKSVGMTKVKVKGEDSSSTFEVEKRRR 237
Db 167 FSVRRFRDITVLADRLAESYR--GFCIPRPDKSV-----VESQVMQKQEFVEQRRV 217
Qy 238 ALERYLQRTVKHPTLLQDPLDPLRFLES--SELPRAVNTQALS----- 277
Db 218 ALEKYLRLVAHPVIRNSDELKVLQROGKLPLATSDVNASRMLDGAVKLPKOLFEGGG 277
Qy 278 -----GAGILRMVNKAADAVNK-----MTIKMNSDAWFEKQOQFENLDQQLR 321
Db 278 ASSVEVQPGRGGRDRLRMFKELRQSVNSDMGSGKPPVVEEDREFLEKKEKMYDLEQII 337
Qy 322 KLHVSVEALVCHRKELSAANTAAFAKSAAMLGNSDHTAL-----SRA--LSQAEVEEKI 374
Db 338 NASQOAEESLVKAQOQDGMGTGELGLAFIKLTKFENEAEVNSQARANDMKNLATSAVKA 397
Qy 375 DQLHQEAFADFYMFSELLSDYIRLITAAVKGVFDHRMKCHQKWDQAQITL-----LKKR 428
Db 398 SRYTRE--LNSQTVKHLDLHDYLGMLMAVQGAADR-----SSALLTVOTLLSELSSL 449
Qy 429 EAEA-KMVA-----NKPDKIOAKNEIREWEAKVQOQGERDFEQISKTIRKEVGRF 478
Db 450 EAREKLEAVASSKVGCDKSRKIEBELKETIKVTEDSKNVAIREYQIKENNNSEVERL 509
Qy 479 EKERVDFKVIKYLESLVOTQOOLIKYWEAFLPEAK 516
Db 510 DRRERADFLNMKGFEVANGVYAEKIANVTKVAEETR 547

RESULT 9
AAG13369
ID AAG13369 standard; Protein; 571 AA.
XX AC AAG13369;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 12839.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123348.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 05-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.

DE Arabidopsis thaliana protein fragment SEQ ID NO: 12838.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
FN EPI033405-A2.
XX
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126284.
PR 01-APR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130510.
PR 30-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134236.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139432.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

```

PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.2%; Score 397.5; DB 21; Length 572;
Best Local Similarity 25.8%; Pred. No. 1.1e-22;
Matches 149; Conservative .93; Mismatches 215; Indels 121; Gaps 18;

QY 22 EDGEDLTSTVSTLESPPSPSPASUPAEIDISANSNGPKTEVWDDDDREDFAEATEEV 81
DB 26 EDGDPLTKSNVG-DKSNYSRSA-----MSTLFDNRHPSIVVTPADSDPLFAPPSIYS 78
QY 82 SLDSPEREPILESSEPSAPVTPPTTLIAPRIESKMSAPVIF-----DRSREIEEANG 137
DB 79 ESRSPKSPNGDGRVSVLYLEPPS-----YADVIFSPFDDISENGSEDCHS 124
QY 138 DIFD-W-----IEIGVSDPEKVGDMNA-----YMAIRVTTKTSLSMFSKSE 177
DB 125 QSSDSLRSRSPSSLDSDYIKTVSNPQKEQATNSMIPGSGTYITYQITRNLSDYGGSE 184
QY 178 FSVKRRSFDELGLHSLKASYLVHGVYIVPAPAEKSIYGMVKVKGEDSSTEFVEKRRR 237
DB 185 FSVRRRERDVTADRLAESYR--GFCIPPRPKSV-----VESQVMQKEVEQRRV 235
QY 238 ALERYLQRTVKHPTLLQDPLRQLES-SELPRAVNTQALS----- 277
DB 236-ALEKYLRLVAHPVIRNSDELKVLQAGKLPATSTDVASRLMDGAVKLPKQLFGEGGG 295

```

```

QY 278 -----GAGILRMVNKAADAVNK-----MTIKMNSDAWFEKQOQFENLDQQLR 321
DB 296 ASSVEVVQGRGRDFLRMFKELRQSVSNDWGGSRPPVVEEDKEFELEKKEKMYDLEQQII 355
QY 322 KLHVSVEALVCHRKELSANTAFAKSAAMLGNSDHTAL-----SRA--LSQLAEEVEEKI 374
DB 356 NASQQAESLVKAQDMGETMGLGFLAFIKLTKFENEAVFNSQARANDMKNLATSAVKA 415
QY 375 DOLHQEQAFADFMFSELLSDYIRLIAAVKGVFDHMKCKQKWEQAQITL-----LKKR 428
DB 416 SRFYRE-LNSQTVKHLDTLHDYLGMLMAYOGAFADR-----SSALLTVQTLSSLSL 467
QY 429 EAEA-KMVA-----NKPKIQQAQKNEIREWEAKVQOGERDPEQISKITIRKEVGRF 478
DB 468 EAAEKLAVASSKVFGDKSRIKKIEELKETIKVTDSEKNVAIREVEQIKENNMSEVERL 527
QY 479 EKEVRKDFKTVIKYLESVQVQQQLIKYWEAPLPEAK 516
DB 528 DRERRADFLNMKMGFVANGVGYAEKIANVWTKVAEETR 565

RESULT 11
AAG41594
ID AAG41594 standard; Protein; 586 AA.
XX
AC AAG41594;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51771.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132486.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.

```


PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 23-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PR	05-MAR-1999;	99US-0123180;
PR	09-MAR-1999;	99US-0123548;
PR	23-MAR-1999;	99US-0125788;
PR	25-MAR-1999;	99US-0126264;
PR	29-MAR-1999;	99US-0126785;
PR	01-APR-1999;	99US-0127462;
PR	06-APR-1999;	99US-0128234;
PR	08-APR-1999;	99US-0128714;
PR	16-APR-1999;	99US-0129845;
PR	19-APR-1999;	99US-0130077;
PR	21-APR-1999;	99US-0130449;
PR	23-APR-1999;	99US-0130510;
PR	23-APR-1999;	99US-0130891;
PR	30-APR-1999;	99US-0131449;
PR	30-APR-1999;	99US-0132048;
PR	30-APR-1999;	99US-0132407;
PR	04-MAY-1999;	99US-0132484;
PR	05-MAY-1999;	99US-0132485;
PR	06-MAY-1999;	99US-0132488;
PR	06-MAY-1999;	99US-0132488;
PR	11-MAY-1999;	99US-0132863;
PR	11-MAY-1999;	99US-0134256;
PR	14-MAY-1999;	99US-0134218;
PR	14-MAY-1999;	99US-0134219;
PR	14-MAY-1999;	99US-0134221;
PR	14-MAY-1999;	99US-0134370;
PR	18-MAY-1999;	99US-0134768;
PR	18-MAY-1999;	99US-0134941;
PR	20-MAY-1999;	99US-0135124;
PR	21-MAY-1999;	99US-0135353;
PR	24-MAY-1999;	99US-0135629;
PR	25-MAY-1999;	99US-0136021;
PR	27-MAY-1999;	99US-0136392;
PR	28-MAY-1999;	99US-0136782;
PR	01-JUN-1999;	99US-0137222;
PR	03-JUN-1999;	99US-0137528;
PR	04-JUN-1999;	99US-0137502;
PR	07-JUN-1999;	99US-0137724;
PR	08-JUN-1999;	99US-0138094;
PR	10-JUN-1999;	99US-0138540;
PR	10-JUN-1999;	99US-0138847;
PR	14-JUN-1999;	99US-0139119;
PR	16-JUN-1999;	99US-0139452;
PR	16-JUN-1999;	99US-0139453;
PR	17-JUN-1999;	99US-0139492;
PR	17-JUN-1999;	99US-0139454;
PR	18-JUN-1999;	99US-0139455;
PR	18-JUN-1999;	99US-0139456;
PR	18-JUN-1999;	99US-0139457;
PR	18-JUN-1999;	99US-0139458;
PR	18-JUN-1999;	99US-0139459;
PR	18-JUN-1999;	99US-0139460;
PR	18-JUN-1999;	99US-0139461;
PR	18-JUN-1999;	99US-0139462;
PR	18-JUN-1999;	99US-0139463;
PR	18-JUN-1999;	99US-0139750;
PR	18-JUN-1999;	99US-0139763;
PR	21-JUN-1999;	99US-0139817;
PR	22-JUN-1999;	99US-0139899;
PR	23-JUN-1999;	99US-0140353;
PR	23-JUN-1999;	99US-0140354;
PR	24-JUN-1999;	99US-0140695;
PR	28-JUN-1999;	99US-0140823;
PR	29-JUN-1999;	99US-0140991;
PR	30-JUN-1999;	99US-0141287;
PR	01-JUL-1999;	99US-0141842;
PR	01-JUL-1999;	99US-0142154;
PR	02-JUL-1999;	99US-0142055;
PR	06-JUL-1999;	99US-0142390;
PR	09-JUL-1999;	99US-0142803;
PR	09-JUL-1999;	99US-0142920;
PR	12-JUL-1999;	99US-0142977;
PR	13-JUL-1999;	99US-0143542;

QY 455 AKVOGERDEQISRTKKEVGEKERVDEKTVIIKYLESLVOTQOOLIKYW 508
Db 517 DAKNVAIKGYIKENNRSEVERLDREDDFMNMKGFFVYVNOGVYAEKMGNW 570

RESULT 13

AAG41595
ID AAG41595 standard; Protein; 566 AA.

XX AC AAG41595;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 51772.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127452.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129645.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134321.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-01467038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139432.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 13.2%; Score 345.5; DB 21; Length 314;
Best Local Similarity 30.6%; Pred. No. 5.8e-19;
Matches 98; Conservative 56; Mismatches 139; Indels 27; Gaps.

QY 142 IEIGVSDPEKVGDMAYMAYRYTTKTSLSMFKSSEFSVKRRFSDFLGLHSLKASLYLHV 201
DB 13 LSVSVTDPVKLGVGQVAYSIRVITTNLPYOGPEKIIRRYSDFVWLDRULFEKY--K 70
QY 202 GYTVPAPAKSIVGMTKKVKYGKSSSTSEFEVEKRRAALERYLQRTVKHPTLLQDPDLROF 261
DB 71 GITPPLPEKSAV-----EKFRSASEFIENRRAALDIFVNRIALHPHQSEDLRTF 122
QY 262 LESSELP-----RAVNTQAL-SGAGILRW----NKAADAVNKMTIKMNESDAWFEKQQO 312
DB 123 LQADEETMDRFREQTISFKKPADLMQMPRDVQSKVSADVLGKEKPVREETADYEKLKH 182
QY 313 FENLDQOLKRLHVSVALVCHRKELSANTAAFAKSAAMGNSEDHTALSRALSQAEEVE 372
DB 183 IFELNHLTEAQKHAYRLVKRRHELQSGLLDFGKAVKLLGACEGEPT-GKAFSDLGTKE 241
QY 373 KIDQLHOEQAFADFYMFSELSDSYIRLIAAVKGVFDRMKWKOWEDAOITLLKKREA 432
DB 242 LLSIKLQKEAQQVLMFFEEPDKDYRVYVQSİKATIAERGTA FKQHCELSETTNDK----- 296
QY 433 KMWANKPKIQQAKNEIRE 452
DB 297 --LMLTRSDKVGEAEIYRE 314

RESULT 15
AAG27810
ID AAG27810 standard; Protein; 327 AA.
AC XX
AC AAG27810;
XX XX
DT 17-OCT-2000 (first entry)
XX XX
DE DE
DE DE
KW KW
KW KW
KW KW
XX XX
XX OS
XX OS
XX OS
PN EP1033405-A2.
XX XX
PD PD
XX XX
PF PF
PF PF
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.

```

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 13.2%; Score 345.5; DB 21; Length 327;

Best Local Similarity 30.6%; Pred. No. 6.1e-19;

Matches 98; Conservative 56; Mismatches 139; Indels 27; Gaps 7;

Qy 142 IEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSVKRRFSDFLGHSKLASKYLHV 201
Db 26 LSVSVTDPKLVONGVOAYISYRVTKTNLPEYQGPKEIVIRYSDVFWLRLFEKY--K 83
Qy 202 GYIVPPAPEKSIIVGMTKVKVGKEDSSSTFEYKRAALERYLQRTVKHPTLLQDPDLRF 261
Db 84 GIFIPPLPEKSAV-----EKFRSAEFTEMRAALDIFVNRIALHPELQQSEDLRTF 135
Qy 262 LESSELP-----RAVNTQAL-SGAGILRWV-----NKAADAVNKWTKMNESDANPEKQQQ 312
Db 136 LQADEETMDRFRFQETSIKRPADLMQMFQVQKSDAVLGKPKPVEETTADYEKLHY 195
Qy 313 FENLDQQLKRLHVSVEALVCHRKELSANTAFAKSAAMGNSEDTALSRLASQLAEVEE 372
Db 196 IFELNHLTEAQKHAYRLVKHRELQGSLLDFGKAVKLLGAGEPT-GKAFSDLGTKSE 254
Qy 373 KIDQLHQBQAFADFYMFSELLSDYIRLIAAVKGVDFHRMKQKQWEDAQITLLKKREA 432
Db 255 LLSIKLQEAQOVLNMFEEPLKDYVRYVQSIKATIAERGTAQKHQHCSELSETTNDK----- 309
Qy 433 KMWANKPDKIQQAKNRE 452
Db 310 --LMLTRSDKVGAEIEYRE 327

Search completed: March 4, 2003, 14:58:35
Job time : 47.4817 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:52 ; Search time 23.3475 Seconds

(without alignments)
2137.006 Million cell updates/sec

Title: US-09-763-902B-3

Perfect score: 2620

Sequence: 1 MAEREPPLGDKPTDFED.....TQOOLIKYWEAFLPEAKAIA 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1554.5	59.3	522	2 G02522	sorting nexin 1 -
2	803.5	30.7	472	2 T31063	hypothetical prote
3	398.5	15.2	554	2 T48483	hypothetical prote
4	309	11.8	420	2 T08691	hypothetical prote
5	254	9.7	312	2 S6952	hypothetical prote
6	243.5	9.3	401	2 T39046	hypothetical phox
7	228	8.7	423	2 S56808	hypothetical prote
8	215.5	8.2	295	2 T42515	hypothetical prote
9	200	7.6	640	2 S67656	hypothetical prote
10	195.5	7.5	361	2 T26610	hypothetical prote
11	167	6.4	2241	2 T20971	hypothetical prote
12	167	6.4	2261	2 T20978	hypothetical prote
13	164.5	6.3	511	2 S53033	hypothetical prote
14	164.5	6.3	2663	1 S28261	centromere protein
15	163.5	6.2	893	2 G88551	protein M01A8.2 (1
16	163	6.2	549	2 T50295	probable vacuolar
17	159.5	6.1	1208	2 T39068	coiled coil protei
18	158	6.0	944	2 S26710	spindle pole body
19	157.5	6.0	823	2 D86165	protein F15K9.3 (1
20	157.5	6.0	918	2 C29829	unknown protein F1
21	157.5	6.0	1133	2 T22976	hypothetical prote
22	156	6.0	143	2 T40731	probable golgi mem
23	156	6.0	764	2 I51302	myosin heavy chain
24	156	6.0	1010	2 T41077	hypothetical struc
25	154.5	5.9	591	2 S51303	hypothetical prote
26	154.5	5.9	1130	2 T21134	hypothetical prote
27	154.5	5.9	1875	2 S38173	myosin-like protei
28	153.5	5.9	992	2 T46337	hypothetical prote
29	151.5	5.8	1104	2 T49735	related to MDM1 pr

30	151.5	5.8	1119	2 T14321	nuclear matrix con
31	151	5.8	1931	2 A59234	slow myosin heavy
32	149.5	5.7	157	2 T18229	hypothetical prote
33	149.5	5.7	807	2 T00990	hypothetical prote
34	149.5	5.7	832	2 T50985	related to transcr
35	149	5.7	1186	2 AD1300	Smc protein essent
36	149	5.7	1356	2 S32763	kinectin 1 - human
37	148	5.6	779	2 T05990	hypothetical prote
38	147.5	5.6	162	2 T26237	hypothetical prote
39	147.5	5.6	1189	2 A54817	ATPase ScII, chrom
40	147.5	5.6	1939	1 A46762	myosin alpha heavy
41	147	5.6	917	2 S64100	probable membrane
42	147	5.6	946	2 A96748	hypothetical prote
43	146.5	5.6	597	2 S40998	hypothetical prote
44	146.5	5.6	1957	2 A45627	myosin heavy chain
45	146	5.6	1938	1 A40997	myosin heavy chain

ALIGNMENTS

RESULT 1

G02522
sorting nexin 1 - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 15-Sep-2000
C:Accession: G02522
R:Kurten, R.C.; Cadena, D.L.; Gill, G.N.
Science 272, 1008-1010, 1996
A:Title: Enhanced degradation of EGF receptors by a sorting nexin, SNX1.
A:Reference number: H01389; MUID:96208617; PMID:8638121
A:Accession: G02522
A:Status: preliminary; nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-522 <KUR>
A:Cross-references: EMBL:U53225; NID:gl293679; PIDN:AAA98672.1; PID:gl293680
C:Genetics:
A:Gene: GDB:SNX1
A:Cross-references: GDB:1391763; OMIM:601272
A:Map position: 15q22-15q23
C:Superfamily: Caenorhabditis elegans hypothetical protein C05D9.1

Query Match	59.3%	Score 1554.5;	DB 2;	Length 522;
Best Local Similarity	59.6%	Pred. No. 3.5e-78;		
Matches 319;	Conservative 78;	Mismatches 99;	Indels 39;	Gaps 8;
QY	2	AAERPPPL-----GDGKPTDFEDLEDGEDLFTSTVSTLESSSPSE--PASLP	48	
DB	10	ASERLPPPPGLPESEGAAGGSEPEAGSDTEGEDIFTG--AAVSKHQSPKITTSLLP	67	
QY	49	AEDISANGPKPTVVLDDRE--DLFAEATEEVSLSDE--REPILSESPAVTPT	104	
DB	68	I-----NNGSKENGTHEDQDQDFADATVELSLDSTQNNQKVLAK-----	111	
QY	105	PTTLAPRTESKMSAPVIFDRSREIEEANGDIFDIGVSDPEKVGDMNAYMAYRV	164	
DB	112	-TLIXLPPQEAFTNSKP---QPTVEEEEEEEDQDLTVGTDPEKIGDMNAYMAYKV	167	
QY	165	TKTSLSMFSKSEFSYKRFSDFLGLHRSKASKYLHVGYIVPPAPEKSIYVGMTKVKVGE	224	
DB	168	TTQTSPLFRSQFQVAKRFESDFGLYKELSEKHSONGFIVPPPEKSLIGHTKVKVGE	227	
QY	225	DSSTFEVEKRAALERYLQRTVKHPTLLQDPLDROFLESSLPRAVNTQALSGAILRM	284	
DB	228	DSSSAEFLKRAALERYLQRIVNHPTMLQDQDVRFELEKEELPRAVGTQTLSSGALLKM	287	
QY	285	VNKAADVANKMTTKMESDANFEKQQQENLDQOLKRLHVSVEALVCHRKELSAATAAF	344	
DB	288	FNKATDAVSKMTTKMESDIWFEEKLOEVECEQRURKLHVVETLVNHRKELALTAQF	347	
QY	345	AKSAAMLGNSDHTALSRLASQAEVEEKIDQLHQEQAFADFYMFSELLSDYIRLTAAYK	404	
DB	348	AKSLAMLGSSDHTALSRLASQAEVEEKTEQLHQEQANNDFLLAELLSDYIRLLAIVR	407	

QY 405 GVFDHMKCKWKEDAQITLLKKREAEAKMMVANKPKIQAKNEIREWEAKVOOGERDF 464
 DB 408 AAFDQRMETWQQAQATLQKKREAEARLLWANKPDKLQAKDEILLESRTVQYERDF 467
 QY 465 EQLSKTIRKEVGRFEKRVKFTVYIYKLSLVQTOQLKYWEAFLEPAKATA 519
 DB 468 ERISTVVRKEVIREPEKSKDFKNHVIKYLETLLYSQOOLAKYWEAFLEPAKATS 522
 RESULT 2
 T31063
 hypothetical protein C05D9.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 15-Sep-2000
 R:Fulton, L.; Gattung, S.
 submitted to the EMBL Data Library, September 1999
 A:Description: The sequence of C. elegans cosmid C05D9.
 A:Reference number: 220960
 A:Accession: T31063
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-472 <FUL>
 A:Cross-references: EMBL:U64605; PIDN:AAB04589.1
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Map position: X
 A:Introns: 70/2; 115/3; 169/3; 296/3; 455/3
 A:Note: C05D9.1
 C:Superfamily: Caenorhabditis elegans hypothetical protein C05D9.1
 Query Match 30.7%; Score 803.5; DB 2; Length 472;
 Best Local Similarity 35.7%; Pred. No. 5.6e-37;
 Matches 174; Conservative 102; Mismatches 178; Indels 33; Gaps 8;
 QY 47 LPAEDISANSNGPKTEVLLDDDDREDLFAEATEEVSIDSPEREPILSSEPSAPVTPVT 106
 DB 2 LSEEDVHTNSVAGRTAVM---EQDLFSDDCDEINLGNED-----TPSKL 45
 QY 107 TLAPRIESKMSAPVFDERSREIEEANGD-----IFDEIGVSDEPKVGDG 155
 DB 46 RLHTE-EPKSSSPAVIN-SIEDHDQYGVNDNHYAQISPEPALSNEPKVTMREREKRGD 103
 QY 156 MNAYMAYRVTKTS-LSMFSGSKSFSVKKRRSDFGLHSLKASKYLHGYIVPPAPEKSIV 214
 DB 104 MNAYIVYKLETVSGVGYTKQHYETWRFSDFGLHGKIVEKTLAKGIVIPOPPEKSIS 163
 QY 215 GMTKVKVGKEDSSSTEVEKRRALRYLQRTVKHPTLLQDPDLROFLE-SSELPRAVNT 273
 DB 164 ALTKTNSDPAMSREVGIGRQLERYICRLIQHPMRNDCDVRFLTIESDLPKAVQT 223
 QY 274 QALSGAGILRWNVKADAVNKMNIKNESDAMFEKQOQFENLDQOLRLKLVSVVEALVCH 333
 DB 224 AALLSFGVKKIFKNFQVVFSAFHFHEEGRWPEQSQVSDDELALRKLTYTVELVAS 283
 QY 334 RKLSANTAFAKSAAMLGNSEDTLSALSOLAEVEEKIDOLHGOAFAFYMSELL 393
 DB 284 RDMATSGEOLGALSMLAAACEESTLSRALSLLTDVTENYSAVYGKQAEVDNSKFSESI 343
 QY 394 SDYIRLIAAVKGVFDHRMKCKWKEDAQITLLKKRAEAK--MMVANKPKIQAKNEIR 451
 DB 344 YEYIMLYSALKDVGGERVRAWQQAQQTARKROQTKIDLSAGRNRSQDLKGEIE 403
 QY 452 ENWAKVOOGERDFEQSKTIRKEVGRFEKRVKFTVYIYKLSLVQTOQLKYWEAF 511
 DB 404 DTQKMDQLEQHEFIELSKAIREEVARFADRKQDKMMLVEYNESTMHTHELLHWK 463
 QY 512 LPEAKAI 518
 DB 464 EPEANNI 470

RESULT 3
 T48483
 hypothetical protein T28J14.60 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.;
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: 224493
 A:Accession: T48483
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-554 <BEV>
 A:Cross-references: EMBL:AL163652
 A:Experimental source: cultivar Columbia; BAC clone T28J14
 C:Genetics:
 A:Map position: 5
 A:Introns: 342/3; 412/3; 499/3; 528/3
 A:Note: T28J14.60
 Query Match 15.2%; Score 398.5; DB 2; Length 554;
 Best Local Similarity 26.0%; Pred. No. 1.2e-14;
 Matches 150; Conservative 92; Mismatches 215; Indels 121; Gaps 18;
 QY 22 EDGEDLFTSTVSTLESSPSPPEASLPALPAEDISANSNGPKTEVVLDDDDREDLFAEATEEV 81
 DB 8 EDGDPLTKSNVNG-DKSNSTRSA-----MSTLFSRHPISIVTVPADSDPLFAPPSSYIS 60
 QY 82 SLOSPEREPILSSEPSAPVTPVTTLIAPRIESKMSAPVIF---DRGREETEEBANG 137
 DB 61 ESRSPKPGNGDRVSSYLEPPS-----YADVIFSPFDDISEINGSEDGHS 106
 QY 138 DIFD-----IEIGVSDEPKVGDGMA-----YMAVRYTTKTSLSMFSKSE 177
 DB 107 QSSDSLRSRSPSSLSDDYIKITVSNPQEQEATNSMIPGGSTYITTYQITRTNLSYDGSSE 166
 QY 178 FSVKRRSDFGLHSLKASKYLHGYIVPPAPEKSIVGMTKVKVGKEDSSSTEVEKRRRA 237
 DB 167 FSVRRRPRDITVLADRLAESYR--GFCIPPRPKSI-----VESQVMQKEVEQRRV 217
 QY 238 ALERYLQRTVKHPTLLQDPDLROFLES-SELPRAVNTQALS-----277
 DB 218 ALEKYLRLVAHPVIRNSDELKVFLOAQGLPLATSTDVASRMLDGAVKLPKQLFGGGGG 277
 QY 278 -----GAGILRWNVKADAVNK-----MTIKNESDAMFEKQOQFENLDQOLRL 321
 DB 278 ASSVEVYVQPGRGGRDFLRNFKELRQSVSDWGGSKPPVVEEDKEFLKKEKMYDLEQII 337
 QY 322 KLHVSVEALVCHKEKLSANTAAFAKSAAMLGNSEDTAL-----SRA--LSQLAEEVKI 374
 DB 338 NASQQAESLVKAAQDDMGETMGLGELAFIKLTKEFNEEAFFNSORARANDMKNLATS AVKA 397
 QY 375 DQLHQOAFADFYMFSELLSDYIRLIAAVKGVFDHRMKCKWKEDAQITL-----LKKR 428
 DB 398 SRYRE-LNSQTVKHLDTLHDYLGMLMAVQGAADR-----SSALLTVQTLSELSSL 449
 QY 429 EABA-KMVA-----NPKDIQAKNEIREWEAKVOOGERDFEQISKTIRKEVGRF 478
 DB 450 EARAELKLVASSVFGDKSRIKKIELKETIKVTEDSKNVAIREVEQIKENWNSEVERL 509
 QY 479 EKERVDFKTVIYKLSLVQTOQLKYWEAFLEPAK 516
 DB 510 DRERRADFLNMKGFFVANQVGYAEKIANVWTKVAEETR 547
 RESULT 4
 T08691
 hypothetical protein DKFZp564F052.1 - human (fragment)
 C:Species: Homo sapiens (nan)
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
 R:Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Cassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16470

A:Accession: T08691

A:Molecule type: mRNA

A:Residues: 1-420 <DUE>

A:Cross-references: EMBL:AL049989

A:Experimental source: fetal brain; clone DKFp564F052

C:Genetics:

A:Note: DKFp564F052.1

Query Match 11.8%; Score 309; DB 2; Length 420;

Best Local Similarity 22.9%; Pred. No. 7e-10;

Matches 113; Conservative 85; Mismatches 171; Indels 124; Gaps 17;

QY 53 SANSNGPKPTVEV-LDDRDLEFAEATEEVSLSPEPILSEPSPAVTPVPTTLIAP 111

DB 3 SSGSALLQAEVLDLEDEDEDFVSKDASLDMSFPMV-----PTSPLSM-- 51

QY 112 RIESKSNAPVIFDRSEETEEANGDIFDIETGVSDPEKVGDMNAYRYVTTKTSL 171

DB 52 -----NQIKFEDEPLDKDLFTVDEPESHVTTIETFIYRIITKTSRG 94

QY 172 MFSKSEFSVRRSDFGLGHSKLASKYLHVGYIVPPAPEKSIVGMTKVKGEDSSSTEF 231

DB 95 EFDSSFEVRRYQDFLWLGKL--EEAHPTLIIPPLPEKFIV-----KGIVERENDDF 146

QY 232 VEKRRALERYLQTVKHPDLLQDPDLRQFLESSELPRAVNTQALSGAGILRMVNKAADA 291

DB 147 IETRRKALHFKELARIADHPHTLTNEDFKIELTAQW--ELSSHKKQGPGLLSRTGTQVRA 204

QY 292 VNKMT-----IKMNSDAWFEKQOQFENLDQLRK-----LHV-- 325

DB 205 VASMRGVKNRPPEFEMNNFIELFSQKINLIDKISRIYKEBEYDEMKEYGPIHLW 264

QY 326 --SVEALVCHRKELSANTAFAKSAAMLGNSDHTALSRLASQLAEVEEKIDQLHQEQAF 383

DB 265 SASSEDLVDTLKVASCIDRCCK-----ATEKRMGSLSEA--LLPVVH----- 305

QY 384 ADYMFSELLSDYIRLTAAYGVDFHRMKCWKQWEDAQITLLKKREAEAKM-----VANKP 440

DB 306 -EYVLYSEM-----MGVNRKRDQI-QAELDSKVEVLYTKKADTLLPEEIGKLE 353

QY 441 DKTOQAKNEIR-EWEAKVQOGERDFEQISKTIRKEVGRFEKERVKDFKTVIKYLESIVQ 499

DB 354 DKYGCANNALKADWE-----RWKQNMNDIKLAFTDMAENIH 391

QY 500 TQOOLIKYWEAFL 512

DB 392 YVEQGLATWESLL 404

RESULT 5

S66952

hypothetical protein YOR069w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein Q2929

C:Species: Saccharomyces cerevisiae

C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 06-Feb-1998

C:Accession: S66952

R:Bohn, C.; Bolotin-Fukuhara, M.; Daignan-Fornier, B.; Dang, D.V.; Valens, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66952

A:Accession: S66952

A:Molecule type: cDNA

A:Residues: 1-312 <BOH>

A:Cross-references: EMBL:274977; NID:g1420218; PID:e252342; PID:g1420219; MIPS:YOR069w

A:Experimental source: strain S288c

C:Genetics:

A:Gene: SGD:VPS5

A:Cross-references: SGD:S0005595; MIPS:YOR069w

A:Map position: 15R

Query Match

Best Local Similarity 9.7%; Score 254; DB 2; Length 312;

Matches 74; Conservative 69; Mismatches 127; Indels 40; Gaps 6;

QY 239 LERYLQTVKHPDLLQDPDLRQFLESSELPRAVNTQA--LSGAGILR----- 283

DB 1 MESLKKIKQDPVQLQDKDFLLFTSDFSSEKSKRAFLTGAINDSNDLSEVRISIEQ 60

QY 284 -----MVNKAADAVNK--MTI-----KMNSDAWFEKQOQFENLDQLRK 324

DB 61 LGAEDAAAEVLKNGGIDAESHKGFMSISFSSLPKYNEADFEFFIEKKQKIDLEDNKKLS 120

QY 325 VSYEALVCHRKELSANTAFA--KSAAMLGNSDHTALSRLASQLAEVEEKIDQLHQEQ 381

DB 121 KSLMEVDTSRNTLAASSTEESSMVETLASLVNSEPNSL--LNNFADVHKSIKSLERS 177

QY 382 APADFYMFSELLSDYIRLTAAYGVDFHRMKCWKQWEDAQITLLKKREAEAKMVAANKPD 441

DB 178 SLOETITMGVYMLDYYIRKSLASVKAIFNQRSLGYFLVVVIENDMKNKHSQGLKLGQNTHSE 237

QY 442 KIOQAKNEIREWAKVQOGERDFEQISKTIRKEVGRFEKERVKDFKTVIKYLESIVQ 501

DB 238 KFEMRKEFOTLERRYNLTKKQWAVGDKIKDEFQGFSTDKIREFRNGMISLEAIESQ 297

QY 502 QOLIKYWEAF 511

DB 298 KECIELMETF 307

RESULT 6

T39046

hypothetical phoX domain containing protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39046

R:Gentiles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, September 1997

A:Reference number: Z21797

A:Accession: T39046

A:status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-401 <GEN>

A:Cross-references: EMBL:298981; PID:CA01735.1; GSPDB:GN000066; SPDB:SPAC6F6.12

A:Experimental source: strain 972h-; cosmid c6F6

C:Genetics:

A:Gene: SPDB:SPAC6F6.12

A:Map position: 1

A:introns: 44/3; 90/1

Query Match 9.3%; Score 243.5; DB 2; Length 401;

Best Local Similarity 23.5%; Pred. No. 2.6e-06;

Matches 97; Conservative 77; Mismatches 146; Indels 93; Gaps 15;

QY 131 IEERANGDIFDIETGVSDPEKVGDG--WNAFYAVYVTKTSLSPKSEFSVKRRFSDFLG 189

DB 7 LDEPSTNSTHFLQCLVTEPRKELQSGSRDTHVYLIITKTNLSIPTRAECKVRRFSDFVK 66

QY 190 LHSKASKYLHVGYIVPAPEKSIVGMTKVKGEDSSSTFEVEKRRALERYLQRTVKH 249

DB 67 LQELSR--WNEDCVPLPAKH-----KLEYIKGGRFSDFNINRAKLLNRYITRCLH 119

QY 250 PTLQDPDLRQFLESSE-----LPRAVNTQALSGAGILRMVNKAADAVNKMTKMN 300

DB 120 PVLHQSPHFIATLENPNWNVVRFIOPKLNNTSKL-----DEISDLSLNAFSLK 170

QY 301 ESDAWFEKQOQFENLDQLRKLVHVSVEALVCHRKELSANTAFAKSAAMLGNSDHTAL 360

DB 171 BEPTFDIQRDHVQOQFMFGISNLEGSQKLLRLEKAL-----DYEDV 214

QY 361 SRLASQLAEVEEKID-----QLHQEQAFADFYMFSELLSDYIRLTAAYKVGV--FDHRMK 412

DB 215 SIOPDLRLASLDQALVDPIESIQNALQQTGTGEYANLTKLT---LLDTIKDVESTYASHLK 271

QY 413 CWQKWEDAQITLLKKREAEAKMVAANKPDKIQQAKNIREWEA-----KVQOGERDFE 465

DB 272 -----ELLARR-----DQKQDVEALQEYSAKLSLRLDKISSGGSGNF 309

hypothetical protein - fission yeast (Schizosaccharomyces pombe)
C5Species: Schizosaccharomyces pombe

QY 129 ETEEEA-----NGDIFDIE-----IGVSDPE 150
Db 108 EYEDSEGLPLNQSNRETCTSLSGSINSNGTSEASEPSVNRKKSARIHILEAK 167
QY 151 KYGDGM-NAYMAYRVTTKTSLSMFSKSEFSVRRFSDFLGLHSLKASKYLHVGYIYPPAP 209
Db 168 RVSEGGGRAYIAVIOFENS-----TVQRRYSDFSLRSILIR--LFPMTLIPP 216
QY 210 E-----KSIVG-WTKVVKVKESSSTE-----FVKKRRAALERYLO 244
Db 217 EKQSIKNYKSGITSSSKYLLPSEGSGVDLSLSVIHASVNNSDKLRIRMLTEFLN 276
QY 245 RTVKHPTLLQDPLROFLES-----SELPRV-----NFQ----- 274
Db 277 KLTNEIETKTSITDLPNNHNWHEFNSSSTFSLPKSILQCNPLDPTNTRIHAML 336
QY 275 ALSGAGILRMVNRKADAVNMKTIKMNSDAMFEKQOQFEN-LDOOL-----RKLHVSVEA 329
Db 337 PIPGSSQLLNKESND-KKMDKERSKSTNIEQDYKQYENLLDNGIYKYNRTTITYHD 395
QY 330 LVCHRKELSANTAFAKSAAMLGN-SEDTALSRALS-QLAEVEEKIDQLHQEQAFADFY 397
Db 396 LKSDYNEIGEVFAQFAHQVQAGELAEQLSYLSNAPSGSSISLEKLVGRU-----YY 447
QY 388 MSELLSDYIRLIAAVKGVDFHRMKQWKWEDAQITLLKKREAEAKMVAN----- 438
Db 448 NINEPLNESVHMATSARELIKRYKLYLQNEIMIKSLNSKRAQLEKLEAONNEYKVDKI 507
QY 439 -----KPD-----KIOQA----- 446
Db 508 IDNEMSKSHITILERNPNNNTSGGKYGKLFNGFNKLASWVKDSVKYQETDPTHTASINL 567
QY 447 KNEIRWEAKVQOGERDFOISKTIRK-EVGRFEKRVKDFTVIYKLYLESVTOQOQLI 505
Db 568 KKEIQLESLEVTENDLEVISKIKNDQLPFKSKEREVDLSILAHYSKRYMENYARQNL 627
QY 506 KYWE 509
Db 628 EIWK 631

RESULT 10
T26610
hypothetical protein Y37A1B.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26610
R:McMurray, A.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z20245
A:Accession: T26610
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-361 <WIL>
A:Cross-references: EMBL:AL023835; PIDN:CAA19488.1; GSPDB:GN00022; CESP:Y37A1B.3
A:Experimental source: clone Y37A1B
C:Genetics:
A:Gene: CESP:Y37A1B.3
A:Map position: 4
A:Introns: 47/3; 161/2; 196/1; 222/1; 277/3; 350/3

Query Match 7.5%; Score 195.5; DB 2; Length 361;
Best Local Similarity 22.4%; Pred. No. 0.00098;
Matches 85; Conservative 68; Mismatches 150; Indels 77; Gaps 15;

QY 146 VSDPEKVG--DGMNAYMAYRVTTKTSLSMFSKSEFSVRRFSDFLGLHSLKASKYLHVGY 203
Db 19 VDKPKESKLGKLSKIATISIT-----SSLTIQORVSRIRYHFDWLHEQLSAKYVLIP- 72
QY 204 IVPAPPEKSVIGTKVVKVKESSSTFEFEKRAALERYIQRTVKHPTLLQDPLRQFLE 263
Db 73 -IPPLEKQVAGRYE-----EDLDHRKHILQLQVWVKICRHPVLSQSEVWLHFTS 121

QY 264 SSE-----LPRAVNTQALSGAGILRMVNRKADAVNMKTIKMNSDAMFEKQOQFENL 316
Db 122 CTDEKDKMKGKRAEYIGGA-----FLNCITVPHQPLDP-----NNV 161
QY 317 DOOL-----RKLHVSVEALVCHRKELSANTAFAKSAAMLGNSEDHTALSRALSOLAEVEE 372
Db 162 DMQVERFQSVKTSSEAMRMQERMNFQKFA--GPVKQNMKGSAFTLQOQSEIDE 219
QY 373 KIDQLHQEQAFADFYMFSELLSDYIRLIAAVKGVDFHRMKQWKWEDAQITLLKKREAEA 432
Db 220 TVASRRUTEALA--YTASE-----YHEIGQVFDATHTK-----NDMEPVL--ENLYSV 262
QY 433 KMMVANKPKIQ--QAKNEIREWEAKVQOGERD-----FEQISKTIKEVGRFEKRVK 484
Db 263 KGTQVQNPDIQVHQAQVQKFRDSEGRSLSSAEAKMKQRIDAMSYTIVIAEVQHQTAEKVE 322
QY 485 DFKTVIIRKYLESLVQTOQOL 504
Db 323 DMKSTMGTYLKKQAMFYQEV 342

RESULT 11
T20971
hypothetical protein F15D3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20971
R:White, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19353
A:Accession: T20971
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2241 <WIL>
A:Cross-references: EMBL:Z81063; PIDN:CAB02951.1; GSPDB:GN00019; CESP:F15D3.1
A:Experimental source: clone F15D3
C:Genetics:
A:Gene: CESP:F15D3.1
A:Map position: 1
A:Introns: 24/1; 130/3; 172/2; 194/2; 206/1; 235/3; 297/3; 438/2; 497/3; 601/2; 737/3
86/3; 2019/3; 2044/3

Query Match 6.4%; Score 167; DB 2; Length 2241;
Best Local Similarity 21.3%; Pred. No. 0.43;
Matches 111; Conservative 86; Mismatches 181; Indels 142; Gaps 23;

QY 91 ILSSEPSA-----VTPVTPTTLIAPRIESKMSAPVIFDRSREIEEANGDIFDIEI 144
Db 257 MLSSQPTSSSSALQIPQTP-----PTAHQAM-----LDRGK-SFEQSAEAGEVRSRK- 304
QY 145 GVSDEPKVGDGMA-----YMAVYRVTTKTSLSMFSKSEF-SVKRRFS 185
Db 305 SSSSSQKSGKSKKARREQLAEFKSCIEQVLTWLLAEDELITLTQMPRVELASVRSQFS 364
QY 186 DFLGLHSLKASKYLHVGYIV-----PPAPEKSVIGTKVVKVKESSSTFEVEKR 235
Db 365 DFESFMSLTSDQTVGVRLLRQMLSNKSESEEEKESICANLHLVNTWEALREQMOE 424
QY 236 RAALERYLQRTVKHPTLLQDPLD-----RQFLESSELP-----RAVNTQAL----- 276
Db 425 QAVLQOOIH-----LQOQSELDTISQWLDAAELIESFGPLAODSSQALRQIELHTKF 477
QY 277 -----SGAGILRMVNRKAA-----DAVNKMTIKMNSDAMFEKQOQFENL 316
Db 478 QOKLNDFOETIDKLESFVAVVDEENDASVATLEDALSAVSVRWGHVCEAEKRAKLDGL 537
QY 317 DOOLRKLHVSVEALVCHRKELSANTAFAKSAAMLGNSEDHTALSRALSOLAEVEEKIDQ 376
Db 538 ADLLDKTNEVFENLSGLAERENELMTGLKSAHLENEEQ---VAQVRRLOKTEEQLEQ 594
QY 377 LHQEQAFADFYMFSELLSDYIRLI-----AAVGVDFHRMKQWKWEDAQITLLKKREAEA 424

Db 595 EH-----ASFVRLSQLSCLVGRLLDSDNGAANAANAVRLSLDSITORDNVLNVARIEEHKTL 649
Qy 425 LKREAEAKMMVANKPKDIOAKNEIREWAKVQ-----QGERDFEQ-----ISK- 469
Db 650 VKSGKADVK-----QVQSQNEQKQEPASSGLSTDTGEQKQNLVDKFLHLHISKL 701
Qy 470 -----TIRKEVGRFEKRVK-DFTKTVIKLYLESIVTQQQ 503
Db 702 SHELEPLQDWSEKFEVSRKKDDIRKMNTCQEKLIQKEQ 741
RESULT 12
T20978
hypothetical protein F15D3.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20978
R:White, S.
submitted to the EMBL Data Library, October 1996.
A:Reference number: 219353
A:Accession: T20978
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2261 <WTL>
A:Cross-references: EMBL:Z81063; PIDN:CA802958.1; GSPDB:GN00019; CESP:F15D3.9
A:Experimental source: clone F15D3
C:Genetics:
A:Gene: CESP:F15D3.9
A:Map position: 1
A:Introns: 24/1; 130/3; 172/2; 211/1; 240/3; 302/3; 443/2; 502/3; 606/2; 742/3; 828/2; 9064/3
Query Match 6.4%; Score 167; DB 2; Length 2261;
Best Local Similarity 21.3%; Pred. No. 0.43;
Matches 111; Conservative 86; Mismatches 181; Indels 142; Gaps 23;
Qy 91 ILSPPSPA-----VTPVPTTLTAPRIESKMSAPVIFDRSREIEEANGDIFDEI 144
Db 262 MLSSQPTSSSSALQIPPTP-----PTAHQAM-----LDRGK-SFEQSAEGEVRSRK- 309
Qy 145 GVSDEKVGCDGNA-----YMYRVTTKLSLMSFKSEF-SVKRRFS 185
Db 310 SSSSQSKSKKARREQLAEFKSCIEQVLWLLAEDELTLTQMPKVELASVRQFS 369
Qy 186 DFLGLHSLASKLVHGYIV-----PPAPEKSIIVGTMKVKGKEDSSSTFEVEKR 235
Db 370 DFEFSSSLTDSQDTVGRVLLRGQMLSNKSESEEEKESIGANHLVNTWREALREQAWQE 429
Qy 236 RAALERYLQRTVKHPTLLQDPDI---RPFLESSELP-----RAVNTQAL----- 276
Db 430 QAVLQOOIH-----LQOSELDTISQWLDAAELIESFGPLAADSSQALRQIELHTKF 482
Qy 277 -----SGAGILRMVAKA-----DAVNMKTIKNESDANFEKQOQFENL 316
Db 483 QOKLNDFOETIDKLESFVAVVENDASVATLEDALSAVSVRMGHVCEWAEKRAKLDGL 542
Qy 317 DQOLRLHVSVEALVCHRLKLSANTAFAKSAAMLGNSDHTALSRAQLAEVEKIDQ 376
Db 543 ADLDDTNEVFENLSGLAERENELMTGLKSAHLENEQ-----VAQVRRLOKTEOLEQ 599
Qy 377 LHQEQAFADYMFSELLSDYIRLI-----AAVKGVFDRHMKCKQKWD-----AQI-----TL 424
Db 600 EH-----ASFVRLSQLSCLVGRLLDSDNGAANAANAVRLSLDSITORDNVLNVARIEEHKTL 654
Qy 425 LKREAEAKMMVANKPKDIOAKNEIREWAKVQ-----QGERDFEQ-----ISK- 469
Db 655 VKSGKADVK-----QVQSQNEQKQEPASSGLSTDTGEQKQNLVDKFLHLHISKL 706
Qy 470 -----TIRKEVGRFEKRVK-DFTKTVIKLYLESIVTQQQ 503
Db 707 SHELEPLQDWSEKFEVSRKKDDIRKMNTCQEKLIQKEQ 746

RESULT 13
S53033
MVPI protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YMR070.06; protein YMR004W
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 30-Apr-1999
C:Accession: S53033; S50223
R:Devlin, K.; Churcher, C.M.
submitted to the EMBL Data Library, March 1995
A:Reference number: S53028
A:Accession: S53033
A:Molecule type: DNA
A:Residues: 1-511 <DEV>
A:Cross-references: EMBL:Z48613; MID:g728645; PID:g728651; MIPS:YMR004W
A:Experimental source: strain AB972
R:Ekema, K.; Stevens, T.H.
submitted to the EMBL Data Library, October 1994.
A:Description: The Saccharomyces cerevisiae MVPI gene interacts with VPS1 and is requ
A:Reference number: S50223
A:Accession: S50223
A:Molecule type: DNA
A:Residues: 1-84, 'N', 86-511 <EKE>
A:Cross-references: EMBL:U16137; NID:g562120; PID:g562121
C:Genetics:
A:Gene: SGD:MVPI
A:Cross-references: SGD:S0004606; MIPS:YMR004W
A:Map position: 13R
Query Match 6.3%; Score 164.5; DB 2; Length 511;
Best Local Similarity 21.5%; Pred. No. 0.079;
Matches 105; Conservative 79; Mismatches 188; Indels 117; Gaps 22;
Qy 32 VSTLESSPSPPEASLPADISANSNGPKPTEVLDDREDLFAATEEV-----SLDSEPR 88
Db 25 VSTTNSPSSL-----NGISGEFNTLNFSTP-----LDTNEEDTGLPTNDVLEESIDDSR 75
Qy 89 EPILSSPSPVPTVPTTLAPRIESKMSAPVIFDR--SREEIEEANGDIF----- 140
Db 76 NPL-----GATGMSQTPNIA-----ANETIDKNDARDQNEIESEADLLDWTNNV 120
Qy 141 -----DIEIGVSDPEKVGDMNAYMAYRVTTKLSLMSFKS-EFSVKRRFSDFLGL 190
Db 121 RKYRPLDADIIIEIEIPREG-LLEKHAHYLVKHLIALPSTSPSEERTVVRYSDFLWL 179
Qy 191 HSKLASKYLHGYIVPPAPEKSIIVGTMKVKGKEDSSSTFEVEKRAALERYLQRTVKHP 250
Db 180 REILLKRYPF--RMIPELPPK-----RIGSONADQL-FLKRRRIGLSRFINLVKHP 228
Qy 251 TLQDPDLRQFLESSELPRAVNTQALS--GAGILRMVAKA--AVNMKTIKNESDAWE 307
Db 229 KLSNDDLVLTF-----TVRTDLTSWRKQATYDTSNEFADKKISQEFMKM-----WKK 276
Qy 308 EKOQFENLDQQLRLKLVSV-----ALVCHRLKLSANTAFAKSAAMLGNSDHTALS 362
Db 277 EFAEQW---NQASCIDTSMELWYRITLLERH-----KRIMQVHERNFEET 322
Qy 363 ALSQAEVEKIDQLHQEQAFADYMFSELLSDYIRLIAAVKGVFDRHMKCKQKWDQA 422
Db 323 LVDNFSEVTPKLYPVQNDTILDNINLSIHKHLETTSSI-----CKQETEIEISG 373
Qy 423 TL-----LKKREAEAKMMVANKPKDIOQ-----AKNEIREWAKVQOGERDFE 465
Db 374 TLPKFKITDILLSLRSLSLFRYKIMAAANNVVELQRHVELNKEKLSMKGKPDVSGAEYD 433
Qy 466 QISKTIKKE 474
Db 434 RIKKIIQKD 442
RESULT 14
S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E

F;80-93/region: nucleotide-binding motif A (P-loop)
F;486-2183/domain: coiled coil #status predicted <COI>
F;92/Binding site: ATP (Lys) #status predicted

Query Match	6.2a	Score	163.5	DB 2	Length	893					
Best Local Similarity	20.1a	Pred.	No.	0.19							
Matches	115	Conservative	94	Mismatches	207	Indels	155	Gaps			
Qy	21	LEDGEDIFTSTVSTLESSPSPASIPAL	DI	SANSGPKPT	TEVVLDDREDLFAE	ATEE	80				
Db	246	IEDNELETPLVET	---	RTMPUL	---	NDLNANSNKNSTTTT	VEPETPKSP	TKKSK	296		
Qy	81	VSLDSPEREPLISEPSPATVPTT	LI	APRITESMS	APVIFDRSREEIEE	ANGDIF	140				
Db	297	M	---	SEKPVVKKPEAPPP	PPFPVKAPSKHQ	---	LMMEQLKASIEAKT	PKPK	348		
Qy	141	DIEICVS	---	---	DPEKVGDMNAY	WAYVTYKTSL	---	SFMSKSEFS	---	VKRRES	185
Db	349	EIKRSVLLPPAPKAPQK	EN	KGEMTETPR	TTTKPLKTVNAK	AKTSPTT	PERQ	---	406		
Qy	186	DFLGLHSKLASKYLHVGIY	IVPPA	PEK	---	---	---	SIVGMTKV	219		
Db	407	---	---	---	---	---	---	---	---	---	---
Qy	220	KVG	---	---	---	---	---	---	---	---	---
Db	457	SKGPFPTSSFAGKLG	OPRTSSSTTS	SAKK	---	---	---	KNPIDEKEKLSR	LQH	505	
Qy	264	SSELPRAV	---	---	NTQALSGAGILRW	NKAADAVN	KNMTIKNES	DAMPEEKQ	QOOF	313	
Db	506	STHAFETLIYVNR	INEDNERKL	---	GNISEY	EKKYSELG	DLKKMLD	EARKKEE	VEQM	563	
Qy	314	ENLDOOLKRLHV	---	SVEAL	---	VCHRKELS	SANTA	FAKS	---	---	---
Db	564	KNSQOQVIRNHAN	AVESLOKTH	ETQIAEKN	KEFERN	PEEERARRE	AEVCAMN	RHKQVYA	623		
Qy	363	ALSO	---	LAEVEEKIDOLH	EOQAFDYMF	SELLSDY	IRLIA	AVKGVDP	HRMKCKQ	WEDAQ	421
Db	624	CLDEKISAEKQC	SOGLNV	DDKV	---	---	---	---	---	---	---
Qy	422	ITLLKRAEAAKMM	---	---	VANKPDK	IQOAKN	BEIREW	EAKVOQ	---	---	---
Db	669	QTALEMKSAMKEL	RQKNLS	LQVDEI	PLKELEISK	WKHKSNEY	KQMLDQ	KINGE	KILV	728	
Qy	465	EOISKTITNKEV	---	GRFEKERV	---	DFKTVI	IKY	493			
Db	729	QOIEDLRKQI	HD	BEKEAMKRS	FDLMQFY	759					

RESULT 15
G88551

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:32 ; Search time 35.5568 Seconds

Amazon, 2009, Enterprise, Second generation, (without alignments) 3007.543 Million cell updates/sec

Title: US-09-763-902B-3

Perfect score: 2620

Sequence: 1 MAAERPPPLGDKPTDFED.....TQQQLIKYWEAFLPEAKAIA 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Minimum DB seq length:	0
Maximum DB seq length:	2000000000

post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 21:★

- ```

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvms:*
16: sp_bacteriap:*
17: sp_archaea:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| 1          | 2561  | 97.7        | 519    | 11    | Q91VZ1 mus musculus |
| 2          | 1514  | 57.8        | 474    | 6     | Q8WNV0 macaca fasc  |
| 3          | 1062  | 40.5        | 458    | 5     | Q9VQ06 drosophila   |
| 4          | 803.5 | 30.7        | 472    | 5     | Q9GV19 caenorhabdi  |
| 5          | 613.5 | 23.4        | 580    | 3     | Q9C2B4 neurospora   |
| 6          | 428.5 | 16.4        | 402    | 10    | Q9FG38 arabidopsi   |
| 7          | 408   | 15.6        | 396    | 10    | Q94CW4 oryza sativ  |
| 8          | 398.5 | 15.2        | 554    | 10    | Q91YQ3 arabidopsi   |
| 9          | 336   | 15.1        | 576    | 3     | Q9C0U7 schizosacch  |
| 10         | 367   | 14.0        | 566    | 10    | Q9FGH8 arabidopsi   |
| 11         | 322.5 | 12.3        | 577    | 10    | Q941X3 oryza sativ  |
| 12         | 261   | 10.0        | 449    | 5     | Q8T0B3 drosophila   |
| 13         | 280   | 9.9         | 417    | 5     | Q9NKJ6 leishmania   |
| 14         | 259   | 9.9         | 228    | 6     | Q9BE62 macaca fasc  |
| 15         | 258   | 9.8         | 336    | 4     | Q8WVD4 homo sapien  |
| 16         | 257.5 | 9.8         | 428    | 5     | Q9V1Q3 drosophila   |

## ALIGNMENTS

## RESULT 1

Q91VZ1 ID Q91VZ1 PRELIMINARY; PRT: 519 AA.

|    |                                                                    |
|----|--------------------------------------------------------------------|
| AC | Q91VZ1;                                                            |
| AD | 01-DEC-2001 (TREMBLrel. 19, Created)                               |
| AE | 01-DEC-2001 (TREMBLrel. 19, Last sequence update)                  |
| AF | 01-JUN-2002 (TREMBLrel. 21, Last annotation update)                |
| AG | RIKEN CDNA 0610030A03 gene.                                        |
| AH | SNX2.                                                              |
| AI | Mus musculus (Mouse).                                              |
| AM | Mus musculus (Mouse).                                              |
| AN | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  |
| AO | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |
| AP | NCBI_TaxID=10090;                                                  |
| AX | NCBI_TaxID=10090;                                                  |
| AY | [1]                                                                |
| BA | EN                                                                 |

Query Match 97.7%; Score 2561; DB 11; Length 519;  
Best Local Similarity 97.9%; Pred. No. 4e-136;  
Matches 508; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

1 MAAEREPPLGDGKPTDFEDLEDGEDLFTSTVSTLESSPSSPEPASLPAEDISANSNGPK 60

[illegible]

Db 1 MAAERPPPLGDKVPTDFEELEDGEDLFTSTVSTLESSPSPASLPAEDISANSNGSK 60

Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:06 11 November 2014

QY 61 PTEVLLDDREDLFAEATEEVSLDSPEREPILSSESPAVTPTTLIAPRIESKMSA 120



```

121 PVIFDRSDEIEEANGDIFDIEIGVSDPEKVGDNAYMAYRVTTTKSLMSKSEFSV 180
181 KRRFDFLHSLKASKYLRHVGIVPPAPEKSIKVGKEDSSSTFEVKKRAALE 240
181 KRRFDFLHSLKASKYLRHVGIVPPAPEKSIKVGKEDSSSTFEVKKRAALE 240
241 RYLQRTVKHPTLLQDPDLROFLFESSSELPRVNTQALSGAGILRMVKNKAADAVNKTIKNN 300
241 RYLQRTVKHPTLLQDPDLROFLFESSSELPRVNTQALSGAGILRMVKNKAADAVNKTIKNN 300
301 ESDANFEKQOQFENLDQOLRLHVSVEALVCHRLKLSANTAFKSAAMLGNSDHTAL 360
301 ESDANFEKQOQFENLDQOLRLHVSVEALVCHRLKLSANTAFKSAAMLGNSDHTAL 360
361 SRALSOLAEVEEKIDQOLHQAFAFADYMFSELSYIRLIAAVKGVDFHRMCKWKWEDA 420
361 SRALSOLAEVEEKIDQOLHQAFAFADYMFSELSYIRLIAAVKGVDFHRMCKWKWEDA 420
421 QITLLKRAEAKMVKANPKDKIOAKNEIREWAKVQOGERDFOISKTIRKEVGRPEK 480
421 QITLLKRETEAKLVANPKDKIOAKNEIREWAKVQOGERDFOISKTIRKEVGRPEK 480
481 ERVKDFKTVIIKYLESLVQTOOOLIKYWEAFLPEAKAIA 519
481 ERVKDFKTVIIKYLESLVQTOOOLIKYWEAFLPEAKAIA 519

RESULT 2
Q8NWV0 PRELIMINARY; PRT: 474 AA.
AC Q8NWV0;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Sorting nexin 1.
GN Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Suzuki Y., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB064995; BAB83537.1;
DR InterPro; IPR001683; PX.
DR InterPro; IPR005329; Sorting_nexin.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF03700; Sorting_nexin; 1.
DR SMART; SM00312; PX; 1.
SQ SEQUENCE 474 AA; 54729 MW; 9E320A2D9782ED0C CRC64;

Query Match 57.8%; Score 1514; DB 6; Length 474;
Best Local Similarity 62.9%; Pred. No. 2.1e-77;
Matches 307; Conservative 70; Mismatches 89; Indels 22; Gaps 6;

QY 36 ESSPSSPEPAS--LPADISANSNGPKPEVVLDDRE--DLFAATEEVSLSDPEREPI 91
DB 5 ESKHQPKRTTSLPI-----NNGSKENGIEHQEQEODLFADATVLSLNTQN--- 55
QY 92 LSSESPAVPTVPTTLIAPRIEBSKMSAPVIFDRGRIIEEANGDIFDIEIGVSDPEK 151
DB 56 -NOKKVPKTLISLPQEAP---NSSKHQP-----IYELEEEEQEDQFDLTGVTIDPEK 106
QY 152 VGDGMNAYMAYRVTTTKSLMSKSEFSVKRRFSDFLGLHSLKASKYLRHVGIVTPAPEK 211
DB 107 IGDGMNAYMAYRVTTTSLPLFRSKOFKAVKRRFSDFLGLYKLSERKHSQNGFTVPPPEK 166

```

```

212 STVGMTKVKVGKEDSSSTFEVKKRAALERYLQRTVKHPTLLQDPDLROFLFESSSELPRAY 271
DB 167 SLIGMTKVKVGKEDSSSAEFLEKRAALERYLQRTVNVHPTMLQDPDREFLEKEELPRAY 226
QY 272 NTOALSGAGILRMVKNKAADAVNKTIKNNESDAWFEKQOQFENLDQOLRLHVSVEALV 331
DB 227 GPTQUSGAGLLKMFNKATDAVSKMTIKNNESDINFEKLEQVECEEQRLKHUVETLV 286
QY 332 CHRKLSANTAAFAKSAAMLGNSDHTALSRLSOLAEVEEKIDQOLHQAFAFADYMFSE 391
DB 287 NHRKELAPNTAQFAKSLAMLGSSDNTALSRLSOLAEVEEKIEQLHQBQANNDPELLAE 346
QY 392 LUSDYIRLIAAVKGVDFHRMCKWKWEDAQITLLKRAEAKMVKANPKDKIOAKNEIR 451
DB 347 LUSDYIRLIAVRAAFDQRMKTQWQDAQATLOKKREAEARLLWANKPKDKLOQADEIL 406
QY 452 EWEAKVQOGERDFOISKTIRKEVGRPEKRVKDFKTVIIKYLESLVQTOOOLIKYWEAF 511
DB 407 EWESRVTOERDFOISKTIRKEVGRPEKRVKDFKTVIIKYLESLVQTOOOLIKYWEAF 466
QY 512 LPEAKAIA 519
DB 467 LPEAKAIS 474

RESULT 3
Q9VQ06 PRELIMINARY; PRT: 458 AA.
AC Q9VQ06;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CG2774 protein.
GN CG2774.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jafarizadeh S., Kalish J., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```

|                                                               |                                                                       |
|---------------------------------------------------------------|-----------------------------------------------------------------------|
| RA                                                            | Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,            |
| RA                                                            | Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,     |
| RA                                                            | Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,           |
| RA                                                            | Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,     |
| RA                                                            | Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L., |
| RA                                                            | Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,        |
| RA                                                            | Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;                      |
| RT                                                            | "The genome sequence of <i>Drosophila melanogaster</i> ";             |
| RL                                                            | Science 287:2185-2195(2000).                                          |
| DR                                                            | EMBL; AE003579; AAF51109.1;                                           |
| DR                                                            | FlyBase; FBgn0031534; CG2774.                                         |
| DR                                                            | InterPro; IPR001683; PX.                                              |
| DR                                                            | Pfam; PF00787; PX; 1.                                                 |
| DR                                                            | SMART; SM00312; PX; 1.                                                |
| SQ                                                            | SEQUENCE 458 AA; 52195 MW; 85ABCAALFI349128 CRC64;                    |
| <br>                                                          |                                                                       |
| Query Match 40.5%; Score 1062; DB 5; Length 458;              |                                                                       |
| Best Local Similarity 43.9%; Pred No. 4.8e-52;                |                                                                       |
| Matches 224; Conservative 89; Mismatches 135; Indels 62; Gaps |                                                                       |
| QY                                                            | 17 DFDELDGEDLFTSTVLTLESSPSPPEASLPADIEDSANGPKPTVEVVLDLDDDLFAE 76       |
| DB                                                            | 1 EVESPEHTRDFAVDINNGAATGSEDEEEVPA-----PGSVTLDRNESDLF-- 48             |
| QY                                                            | 77 ATEVSLDSPERPEILSSPSPAVTVPVTTTL--IAPRIESKSMSPAVTFDRSRREEIEE 134     |
| DB                                                            | 49 -----VSALSPPSGIDIHPLQE-----VLTD----- 69                            |
| QY                                                            | 135 ANGDIPIEIGVSDPEKVGDMNAYWAYRVTTKTLSMFSKEEFSVKRRFSDFLGLHSKL 194     |
| DB                                                            | 70 -DGYDF-ISIVSDPQRIGDMGSYLAKVYTKTNIPKFRSEFTLRFRSFDFLGIDHL 127        |
| QY                                                            | 195 ASKYLHVGYVTPPAPEKSIVGMGTKKVKGEDSS-----STEFVEKRRAALERYLORTVKH 249  |
| DB                                                            | 128 VGKYMRLGRIIIPAPSKNIIGTSFKVISPOSEPGPMPQTQEWVEIRRAALERFVHTAQH 187   |
| QY                                                            | 250 PTLQDPDLRQFLIES-SELPRAVNTQALSGAGILRMVNKAADAENVKMTTKMNESDAWFE 308  |
| DB                                                            | 188 PVLRVLDLDFMNFLESQDELPRSIVNTSALSGAGVIRLFNVGETVKNITYKMNDENPWFD 247  |
| QY                                                            | 309 KQOOFENLQOLRKLVSHVSEALVCHREKLEISANTAAFAKSNAMLGNSEDHTALSALSQA 368  |
| DB                                                            | 248 KITEVESLDANLKQLNHAMKSLVTSRELSLITGLVAKSAAMLSTCEEHTGLSRALSNA 307    |
| QY                                                            | 369 EVEEKIDQLHQEQAFADFVNFSELLSDYIRLIAAVKGVDFHRMKCKQKWEDAQITLLKKR 428  |
| DB                                                            | 308 DVEEKIELLRSEQANSDFFLIAEFINDYLGFGAIKCIPHERVKAQFNQYQAMQLSKER 367    |
| QY                                                            | 429 EAEEKMMVANKPKIOAKNEIREWEAKVOQGRODFEQISKTIKREVGFPERKVKDFKT 488     |
| DB                                                            | 368 ENRGREFLANRADKLDAQOEVDWGQVKVQCQQQDFDSAEIKREWERFELTRVVKDFKV 427    |
| QY                                                            | 489 VIKYLESIVTQOOLIKYWEAFLEPAKAI 518                                  |
| DB                                                            | 428 NIIFYEDMAHQOOIVSYWEAFAPFAEI 457                                   |
| <br>                                                          |                                                                       |
| RESULT 4                                                      |                                                                       |
| Q9GYJ9                                                        | PRELIMINARY; PRT; 472 AA.                                             |
| ID                                                            | Q9GYJ9                                                                |
| AC                                                            | Q9GYJ9;                                                               |
| DT                                                            | 01-MAR-2003 (TrEMBLrel. 16, Created)                                  |
| DT                                                            | 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)                     |
| DT                                                            | 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)                   |
| DE                                                            | Hypothetical 54.1 kDa protein.                                        |
| GN                                                            | C05D9.1.                                                              |
| OS                                                            | Caenorhabditis elegans.                                               |
| OC                                                            | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  |
| OC                                                            | Rhabditidae; Peloderinae; Caenorhabditis.                             |
| OX                                                            | NCBI_TaxID=6239;                                                      |
| RN                                                            | [1]                                                                   |
| RC                                                            | SEQUENCE FROM N2.                                                     |
| RP                                                            | STRAIN-BRISTOL N2.                                                    |

```

RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RN Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton L., Gattung S.;
RT "The sequence of C. elegans cosmid C05D9.";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U64605; AAF98602.1;
DR InterPro: IPR001683; PX.
DR Pfam: PF00787; PX; 1.
DR SMART: SM00312; PX; 1.
KW Hypothetical protein.
SQ SEQUENCE 472 AA; 54107 MW; BFFB40F307F2AF60 CRC64;

Query Match 30.7%; Score 803.5; DB 5; Length 472;
Best Local Similarity 35.7%; Pred. No. 1.6e-37;
Matches 174; Conservative 102; Mismatches 178; Indels 33; Gaps

Qy 47 LPAEDISANSNGPKPTBEVLDDRDREDLFAEATEVSLDSPEREPILSSEPSPAVTPVTPT 106
Db 1 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2 LSEEDVHTNSVAGTVAVM--SOPDLFSDCDEINLGNED-----TPPSKL 45

Qy 107 TLIAPIRESKSMGAPVIFDRSREIEBEANGD-----IFDIEICVSDPEKVGDG 155
Db 1 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46 RLHTE-EPRESSPAVIN-SIEDHQYVGNNDHYAQISPEPALSFKYTMREFEKG DG 103

Qy 156 MNAYMAYRVTKTS-LSMFSKSEFVKRRPSDFGLHSLKASKYLHVGIYVPAPEKSIV 214
Db |||| ||| ||| ||| ||| |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 MNAYIVYKLETVSGVGYTKQHYETWRRESDFGLHGLKIVKYLAKGIVIPQPEKSIS 163

Qy 215 GMTKVYKGDSSSTEFVEKRAALERYLQRTVKHPTLLQDPDLQFLE--SSELPRAVNT 273
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
164 ALTKTKTNSOPAMSREVGIOQARQLERYICRLTOHPMRNDCDVRDLFTIESDLPKAVOT 223

Qy 274 QALSAGATLRMVKNAADAVNKMWTIKMESDAWFEKQOQFENLDQOLRLKLVHSEVALVCH 333
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 AALSFGVKIKFNQVVFVKMAFHMEEGDFEQVQSQVDELDALRLKYTYTETLVAS 283

Qy 334 RKELSANTAFAPAKSAAMLGNSEDHIALSRALSOLAEVEEKIDQLHQEQAFADFYMFSELL 393
Db |||| ||| ||| ||| ||| |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| |||
284 RRDMASTGEOLGKALSALACEESTLSRALSSLTDTVENVSAVYGKQAEVDNDSKFSESI 343

Qy 394 SDYRLTAANKGVFDHMKCKQKWEADAQIITLLKKREAEK--MMVANKPKDIOAKNEIR 451
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 YEYIMLSALKDVFGEVRVAMQOQDAQTLAKRKQDKTKIDLSAGGRNERSDQLGEIE 403

Qy 452 EWEAKVOOGERDFEQISGKTIKREKREKVRDFKVIKIKLESIVQTOQLIKYWEAF 511
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 DTQVKMOLEQHFELSKAIEEVAREADADRQDKMKMLVEYMESMIHTHTELLHLWEAF 463

Qy 512 LPEAKAI 518
Db |||
464 EPEANNI 470

RESULT 5
QC2E4 ID Q9C2E4 PRELIMINARY; PRT; 580 AA.
AC Q9C2E4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Related to vacuolar protein sorting-associated protein VP55.

```

```

RESULT 5
Q9C2E4
ID Q9C2E4 PRELIMINARY; PRT: 580 AA.
AC Q9C2E4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Related to vacuolar protein sorting-associated protein VP55

```

GN 966.210.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA German Neurospora genome project;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL513463; CAC28769.2;  
 DR InterPro; IPR001683; PX.  
 DR Pfam; PF00787; PX; 1.  
 DR SMART; SM00312; PX; 1.  
 SQ SEQUENCE 580 AA; 64100 MW; 311F73EF889C7E2F CRC64;

Query Match 23.4%; Score 613.5; DB 3; Length 580;  
 Best Local Similarity 29.6%; Pred. No. 9.9e-27;  
 Matches 166; Conservative 82; Mismatches 224; Indels 89; Gaps 10;

QY 8 PPLGDKPTPEDLEDGEDFTSTVSTLESPPSPSPASPAPEDISANSNGPPTTEVLD 67  
 Db 22 PQSDDTAPADSTAPSTSTLTPSANTSSSRPRLTPRLVAQ-----PTRLEAVK 73  
 QY 68 DREDLEAEATEEVSLSPP----- 86  
 Db 74 DPLGPL--GASTPTATDTPPLAPOPPLKEQLPTLTGSGPNOIGSGARRPGGSPDHR 131  
 QY 87 --EREPILESPPSPVPTPTTLIAPRIEKSMSAPVIFDRSEETEEBANGDIFDIEI 144  
 Db 132 IEDEELYNSSSGPQPPVPPALPSPVPTSMOPSVSI-----EQAAPTPHISV 181  
 QY 145 GVSDEPKVGMGNMAYRYTTKTSLSMFSKSESVKRRFSDFLGLHSLKSLKYLHVYI 204  
 Db 182 G--DPHKVGLTSSHIYVSVTKTSKAYKOPEFEVKRRYRDLFWLYNTLHAN--NPGV 237  
 QY 205 VPPAPEKSIQVMTKVKVKGEDSSSTEFVKRRALERYLQRTVKHPTLLQDPLROFLES 264  
 Db 238 VPPPEKQAVGRFE-----SNFEARRALEKMLNKTAHPHTLQHDADUKLFLES 287  
 QY 265 SELPRAYNTQAL-----SGAGILRMYNKADAVNKMTIKMNSDANFEEKQOQFENLDQ 318  
 Db 288 EAFNIDVKHKEKPHLGESKGVLTFTGISVSGNKFV---EQDDMFHRRVYLDALEN 343  
 QY 319 QLRKLHVSVEALVCHRKELSAFAKSAAMLGNSDHTALSRLSOLAEEVEKIDOLH 378  
 Db 344 QLKGLKAMDSMVAQRKAMAEAGEFSASLHALSTVLSPTLSGLDALSQLTIRDVY 403  
 QY 379 QEQAFADYFMSSELLSDYIRLIAAVKGVDFHMKCWQKQWEDAOITLLKKREAEKMWAN 438  
 Db 404 DRQAQDVLVTGIIIEYIRLIGSVKQAFSQKQAFYSWHSASELQKRASQDKLLRQ 463  
 QY 439 K--PKIQOAKNEIREWAKVQOGERDEQISKTIRKEVGRFERKRVKDFKTVIKVLES 496  
 Db 464 KQOQRLNQVNAEVADEARKVQARLLFEDGRLMRRELDRFEREKVEDFSGVETFLFS 523  
 QY 497 LVQTOQQOLIKYWEAFLPEAKA 517  
 Db 524 AVEAQKELIEKWETFLMQLDA 544

RESULT 6  
 Q9FG38 PRELIMINARY; PRT; 402 AA.  
 AC Q9FG38;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 21, Last annotation update)  
 DE Sorting nexin-like protein.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP02030; BAB10207.1;  
 DR EMBL; AP002544; BAB10207.1; JOINED.  
 DR InterPro; IPR001683; PX.  
 DR Pfam; PF00787; PX; 1.  
 DR SMART; SM00312; PX; 1.  
 SQ SEQUENCE 402 AA; 46523 MW; 42EC6B67942F7D20 CRC64;

Query Match 16.4%; Score 428.5; DB 10; Length 402;  
 Best Local Similarity 29.8%; Pred. No. 1.5e-16;  
 Matches 115; Conservative 76; Mismatches 172; Indels 23; Gaps 8;

QY 142 IEIGVSDPEKVGDMNAYRYTTKTSLSMFSKSESVKRRFSDFLGLHSLKSLKYLHV 201  
 Db 26 LSVSVTDPVKLGNGVQAYISYRVITKTNLPEYOGPEKIVIRRYSDFWLDRDLFEKI--K 83  
 QY 202 GYVPPAPEKSIQVMTKVKVKGEDSSSTEFVKRRALERYLQRTVKHPTLLQDPLROF 261  
 Db 84 GIFPPLPEKSAV-----EKFRSAEFLMRRAALDIFVNRIALHPELQOSEDRLTF 135  
 QY 262 LESEELP-----RAVNTQAL--SGAGILRMV---NKAADAVNKMTIKMNSDANFEEKQOQ 312  
 Db 136 LQADEETMDRFQETSIKPKPADLMQMFQVQSDAVLGRKPKVEETADYEKLKH 195  
 QY 313 FENLDOLRLHVSVEALVCHRKELSAFAKSAAMLGNSDHTALSRLSOLAEEVE 372  
 Db 196 IFELNHLTBAQHAKVRLVRRHRELQSLDQKAVKLGACGEPT--GAFSDLGTKSE 254  
 QY 373 KIDQLHQEQAFADYFMSSELLSDYIRLIAAVKGVDFHMKCW--QKWEDAOITLLKKREAE 431  
 Db 255 LLSIKLQKAEQAQVLMNFEELPKDYRVYVQSIKATIAERGTAFKQHCSELSETTKLK--EIN 312  
 QY 432 AKMWANPKDIOQAKNEIREWAKVQOGERDEQISKTIRKEVGRFERKRVKDFKTVII 491  
 Db 313 LDKMLTRSDKVGAEIYEYRKAESAEATRRFERIVKRMEDIEIVRQEQTEEMGVAFH 372  
 QY 492 KYLESVQTOQOLIKYWEAFLPEAKA 517  
 Db 373 QFAKQARLANSVADAWRSLLPLEA 398

RESULT 7  
 Q94CW4 PRELIMINARY; PRT; 396 AA.  
 AC Q94CW4;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE P0423B08.7 protein.  
 GN P0423B08.7.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoidae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 clone: P0423B08.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.



Matches 128; Conservative 112; Mismatches 241; Indels 98; Gaps 14;

QY 22 EDGEDLTSTVSTLES-SPSSPEASLPAEDISANSNGPKTEVYL 66  
Db 10 GGDHPLKISDVNGDTSNSGYRSAMSTLSNVRDPLSPPT-----VMI 51  
QY 67 DDDREDLFAEATEEVSLSDPEREPILSSPSPAVTPTTTLIAPRIESKMSAPVIFDR 126  
Db 52 PADSDPLAPSSVEDPRSSFSKPI-SSDMS-----YIEPPSY-ADVIFSP 95  
QY 127 SREEIEEANG-----DIFD-----IEIGVSDPERKVD-----GMNAY 159  
Db 96 FOENSDSEINGTENDSLHQSDFSLSRSSSSSDYIKITVSNPQKEQETSNSIVGGNTY 155  
QY 160 MAYRVTTKTSLSMF-SKSEFSVKRRSDGLHLSKLSKYLHVGIVVPAPAKSIVGMK 218  
Db 156 ITYQITRTNLNLPDGGPSEFSVRRRFRDVTTLADRLAETRY--GFCIPPRPKDSV----- 208  
QY 219 VKVGKEDSSSTEFVKRRRAALERYLQRTVKHPTLLQDPLDQFLE--SSELPRAVNTQALS 277  
Db 209 --VESQVMQKQEFVEQRRVALEKYLRLSAHPVIRNSDELKVFLOVQGLPLPMSTDVAS. 266  
QY 278 -----SRA--LSOLAEVEEKIDQLHQAFAFDFYMFSELLSDYIRLIAAVKGVDFHRMKC 413  
Db 267 RMLDGAVKLPKQLFEGGASAVPVTEVGOPARGGROLLRLFLKELQSVSNMDGGSKPPVV 326  
QY 301 ESDAMFEKQOQFENLDQOLKRLHVSVEALVCHRKLSANTAAFAKSAAMLGNSDHTAL 360  
Db 327 EDKFELEKEXMHDLQEQIINASQAQSLVKAQODMGELGSLAFIKLTFENEAV. 386  
QY 361 -----SRA--LSOLAEVEEKIDQLHQAFAFDFYMFSELLSDYIRLIAAVKGVDFHRMKC 413  
Db 387 CNPQRTANDKMLATAAANKASRYRE--LNSQTVKHLDTLHLYLGMMAVOGAFADRSSA 445  
QY 414 -----WQKEDAQITLLKKEAEAKMMYANKP--DKIQAKNEIREWEAKVQOGERDF 464  
Db 446 LLTVQTLSELLSLOTRVEKLEAASSKVEGGDSRIRKIEELKETIKVTEDAKNVAIKGY 505  
QY 465 EQISKIRVEGVFEKERVKDFKTVIKKILESLOVQOQLIKYW 508  
Db 506 ERIKENRSEVERLDRERRADFNMMKMGFYVNOVGVAEKMGNYM 549

RESULT 11

Q941X3 PRELIMINARY; PRT; 577 AA.

ID Q941X3  
AC Q941X3;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE B1088C09.3 protein.  
DE B1088C09.3.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_Taxid=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC  
clone:B1088C09.";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF003734; BAB68097.1;  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR001683; PX.  
DR Pfam; PF00787; PX; 1  
DR SMART; SM00312; PX; 1.  
SQ SEQUENCE 577 AA; 62866 MW; A3EF68B68FCB31D CRC64;

Query Match 12.3%; Score 322.5; DB 10; Length 577;  
Best Local Similarity 22.9%; Pred. No. 2.1e-10;  
Matches 127; Conservative 81; Mismatches 213; Indels 133; Gaps 17;

Matches 128; Conservative 112; Mismatches 241; Indels 98; Gaps 14;

QY 22 EDGEDLTSTVSTLES-SPSSPEASLPAEDISANSNGPKTEVYL 66  
Db 10 GGDHPLKISDVNGDTSNSGYRSAMSTLSNVRDPLSPPT-----VMI 51  
QY 67 DDDREDLFAEATEEVSLSDPEREPILSSPSPAVTPTTTLIAPRIESKMSAPVIFDR 126  
Db 52 PADSDPLAPSSVEDPRSSFSKPI-SSDMS-----YIEPPSY-ADVIFSP 95  
QY 127 SREEIEEANG-----DIFD-----IEIGVSDPERKVD-----GMNAY 159  
Db 96 FOENSDSEINGTENDSLHQSDFSLSRSSSSSDYIKITVSNPQKEQETSNSIVGGNTY 155  
QY 160 MAYRVTTKTSLSMF-SKSEFSVKRRSDGLHLSKLSKYLHVGIVVPAPAKSIVGMK 218  
Db 156 ITYQITRTNLNLPDGGPSEFSVRRRFRDVTTLADRLAETRY--GFCIPPRPKDSV----- 208  
QY 219 VKVGKEDSSSTEFVKRRRAALERYLQRTVKHPTLLQDPLDQFLE--SSELPRAVNTQALS 277  
Db 209 --VESQVMQKQEFVEQRRVALEKYLRLSAHPVIRNSDELKVFLOVQGLPLPMSTDVAS. 266  
QY 278 -----SRA--LSOLAEVEEKIDQLHQAFAFDFYMFSELLSDYIRLIAAVKGVDFHRMKC 413  
Db 267 RMLDGAVKLPKQLFEGGASAVPVTEVGOPARGGROLLRLFLKELQSVSNMDGGSKPPVV 326  
QY 301 ESDAMFEKQOQFENLDQOLKRLHVSVEALVCHRKLSANTAAFAKSAAMLGNSDHTAL 360  
Db 327 EDKFELEKEXMHDLQEQIINASQAQSLVKAQODMGELGSLAFIKLTFENEAV. 386  
QY 361 -----SRA--LSOLAEVEEKIDQLHQAFAFDFYMFSELLSDYIRLIAAVKGVDFHRMKC 413  
Db 387 CNPQRTANDKMLATAAANKASRYRE--LNSQTVKHLDTLHLYLGMMAVOGAFADRSSA 445  
QY 414 -----WQKEDAQITLLKKEAEAKMMYANKP--DKIQAKNEIREWEAKVQOGERDF 464  
Db 446 LLTVQTLSELLSLOTRVEKLEAASSKVEGGDSRIRKIEELKETIKVTEDAKNVAIKGY 505  
QY 465 EQISKIRVEGVFEKERVKDFKTVIKKILESLOVQOQLIKYW 508  
Db 506 ERIKENRSEVERLDRERRADFNMMKMGFYVNOVGVAEKMGNYM 549

RESULT 10

Q9FGH8 PRELIMINARY; PRT; 566 AA.

ID Q9FGH8  
AC Q9FGH8;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Similarity to sorting nexin.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OX eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
RA Tabata S.;  
RL "Structural analysis of Arabidopsis thaliana chromosomes 5. XI.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB025632; BAB10259.1;  
DR InterPro; IPR001683; PX.  
DR Pfam; PF00787; PX; 1  
DR SMART; SM00312; PX; 1.  
SQ SEQUENCE 566 AA; 63201 MW; 83E6363BB9B1079B CRC64;

Query Match 14.0%; Score 367; DB 10; Length 566;  
Best Local Similarity 24.5%; Pred. No. 6.6e-13;  
Matches 143; Conservative 94; Mismatches 217; Indels 130; Gaps 20;

**Query Match**     **10.0%**; Score **261**; DB **5**; Length **449**;  
**Best Local Similarity** **21.3%**; Pred. No. **4.3e-07**;  
**Matches** **114**; Conservative **88**; Mismatches **206**; Indels **128**; Gaps **16**;

**Qy**       **23 DGEDLFTSTVTSLESSSPSPALPAEDISANGNGPKTEWVLDDDRDLFAEATEVS 82**

            |||     | | : |||     | : :     | : : :     | : : :

3 DGDG-----DSNLLNSPSTNGATM-----ETASPTN-----P 30

83 LDSPEREPLISBPSPAPVPTVPTTLLIAPRIEKSMSGAPVIFDRSREIEEANGIDFI 142

31 LATP---PATGGAPAPSGATNGSGSATSPDSSSSAPATPAVLGENALRVE-----I 78

143 EIGVSDPEKVGDMNAYMAYRVTKTSLSMFSEKSEFVKRRFSDFLGHLKSLKYLHV 202

79 SDALSEKEV-----KFTVHTRTTLPFGSKDNVVRHOEEFVWLHDIENDDIYAG 130

203 YIYVPAPEKSIYVGTQKVKVGK-----EDSSSTEFVEKERRAALER-----YL 243

131 YIYPCPRPDDFASREKLQRLGSGEGENWKKEEFKMKKSELEAEYLATFKTYAMHEVFL 190

244 QRTYKHPDLLQDPLROFLESSE-----LPRAVNTQALSGAGILRMVNAADAVNMKTIKM 299

191 RRLASHPVFRVDHLKVFLEYDQDLCAKPR--KKMAIFG-GFVKSLGKTTTDEI--LLSATV 246

300 NESDAWEEKQOQFENLDQOLRLKHSVEALVCHRKELSANTAFAKSAAMLGNSDHHTA 359

247 RDVNDFFENELQFLTEYHGHLREAAALTEKMTQRHKOV-----GDSHQK 290

360 LSRALSO-----AEVEEKIDOLHGOEAFADYMFSELLSDYIRLIAAVK 404

291 ISNALTQLSSTEKENGVETFAVKAETAEIERIKNLNTRVASQDQLGLDRLRYQRDSDAAK 350

405 GVFDHRMKQKQWEDAQITLLKKREAAKMMVANKPKIQOAKN-EIREWEAKVQOGERD 463

351 ALLIRLRCLAAEYAAANRLEKAR-----SKNDVHAAETAQAEACEK 393

464 FEQISKTRKEVGFKEKRVKDFVTIYKIYLESVOTQOOLIKYWEAFLPEAKAIA 519

394 FESMAGCKEELIGFRNRRVAAAFKKSIVSELSEBIKHAQTOYELROSLLALKBEIA 449

RESULT 13

Q9NKJ6 PRELIMINARY; PRT; 417 AA.

ID Q9NKJ6

AC Q9NKJ6;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Sorting nexin 1.

GN SNX1.

OS Leishmania major.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

OX NCBI\_TaxID=5664;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FRIEDLIN;

RA Myler P.J., Sisk E., Cawthra J., Handley F., Vogt C., Robertson L.,

RA McDonagh P., Ivans A., Nguyen D., Munden H., Stuart K.,

RL Submitted (MAY-2000) to the EMBL/genbank/DBJ databases.

RL EMBL; AC013257; AAF69581.1;

DR InterPro; IPR001683; PX.

DR Pfam; PF00787; PX; 1.

DR SMART; SM00312; PX; 1.

SQ SEQUENCE 417 AA; 46558 MW; 718450F7D4DF86C2 CRC64;

Query Match 9.9%; Score 260; DB 5; Length 417;

Best Local Similarity 23.3%; Pred. No. 4.5e-07;

Matches 98; Conservative 77; Mismatches 190; Indels 56; Gaps 11;

Qy 142 IEIGVSDPEKV-GDGM--NAYMAYRVTKTSLSMFSEKSEFVKRRFSDFLGHLKSLKYL 198

Dy 7 ISVKVDVPSQVKGQGTLEMSYIYIPITMLPFGM---SDARFNRRYITDFLRLGOLCATY 63

Qy 199 LHVGVIYVPAPEKSTIVGMTKVKVGK-----EDSSSTEFVEKERRAALERYLQRTVK 248

Dy 64 WYC--IVPIPEKSV---QDKLGLPRWASAKETTASEGDLLEYRISLRFLQRLAY 118

Qy 249 HPTLLQDPLRLQFLESSEL-----PRAVNTQAL-----SGAGILRM- 284

```

Db 119 HPIIGKSDLLQKFTNDNEWQICIRDPVPPRFTIVSSLEETARSWPFSSSAGAGGSG 178
Qy 285 -----VNKAADVNTKMTIKMNSDAWFEKQOOFENLDOOLKRLHVSVEALVCHRKEL 337
Dt 01-MAR-2002 (Tremblrel. 17, Last sequence update)
Db 179 GAAGGGVSGTGAAYQSALTQEPVDEATWRATSEYIGLESNKNMRNLLALVDRHRA 238
Qy 338 SANTAFAKSAAMLGNSDHTALSRALSQAEVEKIDOLHOEQAFADYMFSELSYDI 397
Dt 01-JUN-2002 (Tremblrel. 21, Last annotation update)
Db 239 ASAVSNFAASFGLLANGBDAELGATGVRDCGKRVADYVKHADNSTRVLSTLSPV 298
Qy 398 RLIAAVKGVFDRHMKCKQKQWEDAQITLLKKREAKMVMVANKPKIQAKNEIREWEAKV 457
Dt 01-JUN-2001 (Tremblrel. 17, Created)
Db 299 GYCAAVRETNLHMFSAHOYLNLQ---KGOELQASAMRAQSNQV-QVQSELHFVNEQR 354
Qy 458 QOGERDFEQIKTKRKEVGRKERVKDFKTVIKYLESVOTQOOLKYWEALPEAKA 517
Dt 01-JUN-2001 (Tremblrel. 17, Last sequence update)
Db 355 AHLEEDLGAECTSEEPVLFHENKQYDAKDKMLKFGMLSEFSSESMMQENDALRPMLES 414
Qy 518 I 518
Db 415 L 415

RESULT 14
Q9BE62 PRELIMINARY; PRT; 228 AA.
AC Q9BE62;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 24.8 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_Taxid=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FRONTAL LOBE LEFT;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB056808; BAB39332.1;
DR InterPro; IPR005329; Sorting_nexin.
DR Pfam; PF03700; Sorting_nexin; 1.
KW Hypothetical protein.
SQ SEQUENCE 228 AA; 24827 MW; FEE036C502053BD4 CRC64;

Query Match 9.9%; Score 259; DB 6; Length 228;
Best Local Similarity 32.2%; Pred. No. 2.3e-07;
Matches 76; Conservative 43; Mismatches 75; Indels 42; Gaps 10;

Qy 2 AAEERPPL-----GDGKPTDFEDLEDGDLTSTVSTLESSPSPSPAS--LP 48
Dt 01-JUN-2001 (Tremblrel. 17, Last sequence update)
Db 10 ASERLPFPFPGLEPESEGAAGSEAGSDTEGDIFTG--AAVYSHQSPKRTTSLP 67
Qy 49 AEDISANSNGPKPTVWLDDORE--DLFAATEEVSLSPEPILSSESPSPVPTVPT 106
Dt 01-JUN-2001 (Tremblrel. 17, Last sequence update)
Db 68 I-----NNGSKNGIHEEQDQEPDLFADAFVLSLNTQN---NOKKVPKTLISLP 117
Qy 107 TLIAPIRESKMSAPVIFDRSREIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRVTT 166
Dt 01-JUN-2001 (Tremblrel. 17, Last sequence update)
Db 118 PQEAP--NSSKHQ-----TYEELEEEQEDQFDLTGVTIDPEKIGDMNAYVAYKVT 169
Qy 167 KYSLSNFKSEFS-----VKRRFSDFLGHSLKASKYLHV--GYIVPAPPEKSI 215
Dt 01-JUN-2001 (Tremblrel. 17, Last sequence update)
Db 170 QVSPGGPAEDHLRLGCLYQRRHSDYFVYSPFLFQHLHLYSRWSKSPASPWRSHLG 225

```

RESULT 15  
Q9WVD4

```

ID Q9WVD4 PRELIMINARY; PRT; 336 AA.
AC Q9WVD4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to sorting nexin 7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC018105; AAH18105.1;
DR InterPro; IPR001005; MYB_DNA_binding.
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 336 AA; 39207 MW; AAE0BE2FA7BF4616 CRC64;

Query Match 9.8%; Score 258; DB 4; Length 336;
Best Local Similarity 24.5%; Pred. No. 4.4e-07;
Matches 92; Conservative 60; Mismatches 138; Indels 86; Gaps 11;

Qy 97 SPVTPVPTTTLIAPRIESKMSAPVIFDRSREIEEANGDIFDIEIGVSDPEKVGDM 156
Db 7 SP-WMPTSPISMI-----NOKKFEDEPLKDLFTYVDEPESHVTTI 46
Qy 157 NAYMAYRVTTKTSLSMFSKSEFSVKRRFSDFLGLHSLKASKYLHVGYIVPPAPEKSI 216
Db 47 ETFITYRIITKTSRGEFSDSEFEVRRVQDFLWLGKL--EEAHTLIPLPEKFI-- 102
Qy 217 TKVKVGKEDSSSTFEVKKRAALERYLQRTVHTLQDPLROFLSESELPRAVNTQAL 276
Db 103 -----KGMVERFNDDFIETRRKALHKLNRADHPTLTFFNEDKIFLTAQAA-- 156
Qy 277 SGAGILRMVNKAADAVNKMT-----TKMNSDAWFEKQOOFENLDOOLK-- 322
Db 157 QGPGLSRMGOTVRVAVASSMRGVKNRPEEFEMNNFELFSQKINLIDKISQRIYKEBE 216
Qy 323 -----LHV-----SVEALVCHRKELSAATAAFKSAAMLGNSDHTALSRAL-- 364
Db 217 YFDEMKEYGPIHILWSASEEDLVDTLKVASCIDRCCA-----TEKRMSSLSEALLPV 271
Qy 365 -----SOLAEVEEKIDOLHOEQAFADYMFSELSYDIIRLIAVKGVPDRHMKCKQK 417
Db 272 HEYVLYSEMIMGVMKRRDQIQAE-----LDSKVEVLTYYKKADTDLCLATWESF 319
Qy 418 EDAQITLLKKREAEAK 433
Db 320 LTSQTNLHLEASEDK 335

```

Search completed: March 4, 2003, 15:02:35  
Job time: 38.5568 secs

| Result<br>No. | Query  |       |        | DB | ID               | Description       |
|---------------|--------|-------|--------|----|------------------|-------------------|
|               | Score  | Match | Length |    |                  |                   |
| 1             | 2033   | 77.6  | 451    | 1  | US-08-625-322-4  | Sequence 4, Appli |
| 2             | 1554.5 | 59.3  | 522    | 1  | US-08-625-322-2  | Sequence 2, Appli |
| 3             | 149    | 5.7   | 1939   | 4  | US-09-310-187A-1 | Sequence 1, Appli |
| 4             | 145    | 5.5   | 2101   | 1  | US-08-466-390-4  | Sequence 4, Appli |
| 5             | 145    | 5.5   | 2101   | 1  | US-08-470-950-4  | Sequence 4, Appli |
| 6             | 145    | 5.5   | 2101   | 1  | US-08-467-781-4  | Sequence 4, Appli |
| 7             | 145    | 5.5   | 2101   | 1  | US-08-195-487-4  | Sequence 4, Appli |
| 8             | 145    | 5.5   | 2101   | 2  | US-08-483-924-4  | Sequence 4, Appli |
| 9             | 145    | 5.5   | 2101   | 4  | US-09-452-294-1  | Sequence 1, Appli |
| 10            | 145    | 5.5   | 2101   | 5  | PCR-US93-06160-4 | Sequence 4, Appli |
| 11            | 143.5  | 5.5   | 1886   | 4  | US-08-938-105-3  | Sequence 3, Appli |
| 12            | 134.5  | 5.1   | 976    | 4  | US-09-104-324B-4 | Sequence 4, Appli |
| 13            | 131.5  | 5.0   | 1093   | 5  | PCR-US93-03077-1 | Sequence 1, Appli |
| 14            | 130.5  | 5.0   | 1829   | 4  | US-09-157-420-1  | Sequence 1, Appli |
| 15            | 128.5  | 4.9   | 372    | 1  | US-07-813-584A-3 | Sequence 3, Appli |
| 16            | 128.5  | 4.9   | 372    | 1  | US-08-330-515-3  | Sequence 3, Appli |
| 17            | 128.5  | 4.9   | 1164   | 4  | US-09-457-708-2  | Sequence 2, Appli |
| 18            | 127.5  | 4.9   | 1786   | 4  | US-08-973-463-8  | Sequence 8, Appli |
| 19            | 127.5  | 4.9   | 3248   | 1  | US-08-353-700-1  | Sequence 1, Appli |
| 20            | 127.5  | 4.9   | 3248   | 5  | PCR-US95-16216-1 | Sequence 1, Appli |
| 21            | 126.5  | 4.8   | 619    | 1  | US-08-465-746-2  | Sequence 2, Appli |
| 22            | 126.5  | 4.8   | 619    | 1  | US-08-214-164-2  | Sequence 2, Appli |
| 23            | 126.5  | 4.8   | 619    | 2  | US-08-467-852A-3 | Sequence 3, Appli |
| 24            | 126.5  | 4.8   | 619    | 2  | US-08-466-636-2  | Sequence 2, Appli |
| 25            | 126.5  | 4.8   | 619    | 2  | US-08-247-491A-3 | Sequence 3, Appli |
| 26            | 126.5  | 4.8   | 619    | 2  | US-08-319-795-2  | Sequence 2, Appli |
| 27            | 126.5  | 4.8   | 619    | 2  | US-08-468-985-2  | Sequence 2, Appli |



```

Db 51 MSAPVIFDRSREIEEANGDIFDIEIGVSDPEKVGDMNAYMAYVTTKTSLSMFSKSE 110
QY 178 FSVKRFSDFLGHSKLASKYLHVGVIVPAPPEKSTVGMTKVKGKEDSSSTFEVEKRA 237
Db 111 FSVK-RFTDFLGLHTLPTLYLHVIVFATSRKSLVGMTKVKGKEDSSSTFEVEKRA 169
QY 238 ALERYLQRTVKRPTLLQDPDLQFLESSELPRAVNTQALSGAGILRMVKNKAADAVNKMFI 297
Db 170 ALERYLQRTVKRPTLLQDPDLQFLESSELPRAVNTQALSGAGILRMVKNKAADAVNKMFI 229
QY 298 KKNESDANFEKQOQFENDQDLRLHVSVEALVCHRKLSANTAFKSAAMLGNSDGH 357
Db 230 KKNESDANFEKQOQFENDQDLRLHVSVEALVCHRKLSANTAFKSAAMLGNSDGH 289
QY 358 TALSRLSOLAEVEEKIDOLHQQEAFADYMFSELSLSDYIRLIAVKGVDHRMCKWOKW 417
Db 290 TALSRLSOLAEVEEKIDOLHQQEAFADYMFSELSLSDYIRLIAVKGVDHRMCKWOKW 349
QY 418 EDAQITLLKREAEAKMMVANKPKDIOQAKNEIREWEAKVOQGERDFQISKTIRKEYGR 477
Db 350 EDAQITLLKREAEAKMMVANKPKDIOQAKNEIREWEAKVOQGERDFQISKTIRKEYGR 409
QY 478 FKERVKDFKTVIIKYLESLVOTQOOLIKYWEAFLEPEAKATA 519
Db 410 FKERVKDFKTVIIKYLESLVOTQOOLIKYWEAFLEPEAKATA 451

```

## RESULT 2

```

US-08-625-322-2
; Sequence 2, Application US/08625322
; Patent No. 5804412
; GENERAL INFORMATION:
; APPLICANT: Gill, Gordon N.
; APPLICANT: Kurten, Richard C.
; APPLICANT: Cadena, Deborah L.
; TITLE OF INVENTION: Sorting Nexins and Methods of Using Same
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,322
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 1955
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-625-322-2

```

```

Query Match 59.3%; Score 1554.5; DB 1; Length 522;
Best Local Similarity 59.6%; Pred No. 1.2e-120;
Matches 319; Conservative 99; Indels 39; Gaps 8;

```

```

QY 2 AAERBPPPL-----GDGKPTDFEDLEDGDLTSTVSTLESSESPSE--PASLP 48
Db 10 ASERLPFPFPLEPSEGAAGSEPEAGSDTEGEDIFTG---AAVYKHSQSPKITSLLP 67
QY 49 AEDISANGNGKPTFVLDVDDRE--DLFAEATEEVSLSPE--REPILSSSESPAVTPVT 104
Db 68 I-----NNGSKENGIEHEEQDQDQLFADATVELSLDSTQNNQKVLAK----- 111
QY 105 PTTLAPRIEKSMSAPVIFDRSREIEEANGDIFDIEIGVSDPEKVGDMNAYMAYV 164
Db 112 -TLIXLPPOEATNSSKP--OPTYEELSEEQEDQDFTVGTIDPEKIGDGMNAYVAVKV 167
QY 165 TTKTSLMSFSEFVKRRFSDFLGHSKLASKYLHVGVIVPAPPEKSTVGMTKVKGK 224
Db 168 TTQTSLPFLFRSKQFAVKRRFSDFLGHSKLASKYLHVGVIVPAPPEKSLIGMTKVKGK 227
QY 225 DSSSTFEVEKRAALERYLQRTVKRPTLLQDPDLQFLESSELPRAVNTQALSGAGILRM 284
Db 228 DSSSAEFLKRAALERYLQRTVKRPTLLQDPDLQFLESSELPRAVNTQALSGAGILRM 287
QY 285 VNKAADAVNKMFIKKNESDANFEKQOQFENDQDLRLHVSVEALVCHRKLSANTAF 344
Db 288 FNKATDAVSKMTIKKNESDIWFEKLEQVECEQRLKHAHVETLVNHRKELANTAQF 347
QY 345 AKSAAMLGNSDHTALSRLSOLAEVEEKIDOLHQQEAFADYMFSELSLSDYIRLIAV 404
Db 348 AKSLAMLGSSDNTALSRLSOLAEVEEKIDOLHQQEANGNDFLLAELLSYIRLIAV 407
QY 405 GVDFHRMCKWOKWEDAQITLLKREAEAKMMVANKPKDIOQAKNEIREWEAKVOQGERDF 464
Db 408 AAFDQRMKTQWQDQAQATLQKREAEARLLWANKPKDIOQAKDEILEWESRVTYQERDF 467
QY 465 EQISKTIRKEVGRKERVKDFKTVIIKYLESLVOTQOOLIKYWEAFLEPEAKATA 519
Db 468 ERISTVYVRKEVIRPEKSKDFKNHVIKYLETLLYSQOLAKYWEAFLEPEAKATA 522

```

## RESULT 3

```

US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; FILE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

```

```

Query Match 5.7%; Score 149; DB 4; Length 1939;
Best Local Similarity 18.1%; Pred. No. 0.002;
Matches 99; Conservative 101; Mismatches 175; Indels 172; Gaps 22;
QY 19 EDLEDGEDLTSTVSTLESSESPSEPAFLPAEDISANGPKPTVEVYLDLDDREDLFAEAT 78
Db 1386 EELAEAKKLAQQLQDAEEA-----VEAVNAKCSSLEKTKHKLQNEIEDLMDYDVE 1435
QY 79 EE-----VSLDSPER--EPILSS-----EPSPAVTPVPTTLIAPRIEKSMSAPVI----- 123
Db 1436 RSNAAALDKQKRNFKDILAEKQKVEES-----QSELESSQKREARSLFELKLN 1488
QY 124 -----FDRSREIEEANGDIFDIEIGVSD--PEKVGDMNAYMAYVTTKTSL 170
Db 1489 AYESLEHLETFKRNKNLQEE-----ISDLTEQLGEGGKNVHELEKV----- 1531

```

QY 171 SMFSKSESVKRRFSDPLGLHSLKSLKYLHVGYIVPPAPEKSIKVGMTKVKVGRDSSSTE 230  
Db 1532 -----RKQLEVEK-----LEQSAL-----EAEASLE 1554  
QY 231 FVEKR--RAALE-RYLQRTVKHPTLLQDPDLRPFLESE-----LPRAVNTQALSGAGIL 282  
Db 1555 HEGKILRAOLEFNOIKAEIERKLAEDDEMEQAKRHHQRVDSLOTSLDAETRSRNEVL 1614  
QY 283 RMYNKAADAVNKTIKMNESDANFEKQOQFENLDQOLRLHVSVEALYCHRKELSANTA 342  
Db 1615 RVKKMEGDLNEMEIOLSHANRAAEAKQVKSLSLLKDTQIQDDAVRANDDLKENIA 1674  
QY 343 AFAKSAAMLGNSE-----DITLSRLALS--QLAEEVEEKIDOLHOGQAFADPYMESEL 392  
Db 1675 IVERNNLL--QAELEELRAVVEQTERSKLAELIETSERVQLLHSQNT-----1723  
QY 393 LSDYIRLIAAVKGVDFHRMKCQKWEDAQITLLKRAEAEAKMVMYKPKIQAKNIRE 452  
Db 1724 -----SLINQKKMESDLTQL-----QSEVEEAVQCBRN 1752  
QY 453 WEAKVOGERDFEIOISKIRKEVGRFEKERVKDKPTVIKYLESLVOTQOOLIKYWEAFL 512  
Db 1753 AEEKAKKAITDAAMABELKKE-----ODTSA-----HLERKKNMEQTIKDLQHL 1799  
QY 513 PEAKAIA 519  
Db 1800 DEAEQIA 1806  
RESULT 4  
US-08-466-390-4  
; Sequence 4, Application US/08466390  
; Patent No. 5698562  
; GENERAL INFORMATION:  
; APPLICANT: TOUKATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,390  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-466-390-4

Query Match 5.5%; Score 145; DB 1; Length 2101;  
Best Local Similarity 18.2%; Pred. No. 0.0047;  
Matches 105; Conservative 104; Mismatches 188; Indels 180; Gaps 20;

QY 20 DLEDGEDLTSTVSTLESSPSPASLPAEDISANSNGPKTEVWLLDDDDREDFAEATE 79  
Db 136 DHEDGLNLEDENFLQAKAPVSTCSSTFFEEELSPPHQAK-----RETFLELQK 186  
QY 80 EVSLDSPEREPIILSSSESPAVTPTTLLIAPRIESKMSAPVIFDRSREEIEEEANGDI 139  
Db 187 VASSSSGNN--FLSGSPSPMGDILQT---POFQMRRL-----KQLADERSNRD- 231  
QY 140 FDEIGVSDDEKVGDDGNAYNAVYVTKTSLSMFSEKSEFSVKRRFSDPLGLHSLKSLYL 199  
Db 232 -ELELELAENRKLKLTEDAQIA-----MMQORIDRLALLNEKQAAS-- 271  
QY 200 HVGIVYVPAPEKSIKVGMTKVKVGRDSSSTEFEVKKRAALERYLQRTVKHPTLLQDPDLR 259  
Db 272 -----PLEPKE-----LEELDKNESLTMRL--HETLKQCODLK 303  
QY 260 QFLESSLPRAVNTQALSGAGILRMVNKAA-----DAVNKMTIKMNESDANFEKQOQ 312  
Db 304 --TEKQMDRKINQLSEENGDLSPKLRFAHLQOLQDALNELTEHSHKATQEWLEKQAO 361  
QY 313 FE-----NLDOOLRLKH-----VSVEALVCH 333  
Db 362 LEKLSAALQDKKCLEKNEITLQKLSLEHLSQLQDNPPQEKGEVLGDVLIQLETLKQE 421  
QY 334 RKELSANTAFA-----AKSAMLGNSEDH-----TALSRLALSQAEV 370  
Db 422 AATLAANNTOQARVEMLETERGQOEAKLLAERGHFEEKQOQLSSLLTDLOSSISLSQA 481  
QY 371 EEKIDQLHQEQAFADFYMFSELLSDYIRLIAAVKGVDFHRMKCQKWEDAQITLLKREA 430  
Db 482 KEELEQASQAHGABRLTAQVASLTSELTNLNATI-----QQOQDELAKLQOAK 529  
QY 431 EAKMVMYKPKIQOQ-----KNEIREWEAKVQOGERDFEIOISKIRKEVGR----- 477  
Db 530 EKQAOQLAQTLOOQEQASOGLRHQVEQLSSSLKQEQOLKEVAE--KQEAOTRODHAQOLAT 587  
QY 478 --FEKERVKDEKTVIIKYLESLVQ-----TOOQL 504  
Db 588 AAEREASLRERDAALKOLEALEKAKEAKLEILOQOOL 624  
RESULT 5  
US-08-470-950-4  
; Sequence 4, Application US/08470950  
; Patent No. 5698439  
; GENERAL INFORMATION:  
; APPLICANT: TOUKATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,950  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000  
 TELEFAX: (617) 248-7100  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2101 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-470-950-4

Query Match 5.5%; Score 145; DB 1; Length 2101;  
 Best Local Similarity 18.2%; Pred. No. 0.0047;  
 Matches 105; Conservative 104; Mismatches 188; Indels 180; Gaps 20;

```

QY 20 DLEDGDLFTSTVSTLESSPSPPEASLPADISANSNGPKPTEVLLDDDDREDFAEATE 79
DB 136 DHDGGLNLEDLENFLQKAPVPTCSSTPEELSPSHQAK-----RETRFLELQK 186
QY 80 EVSLDSPEREPIISSEPSPAVPTPTTLIAPRIEKSMSAPVIFDRSREIEEANGDI 139
DB 187 VASSSSGNN--FLSGSPASPMGDILOT---POQMRRL-----KKQLADERSNRD- 231
QY 140 FDEIGVSDPEKVGDCGNMAYMAYRVTTKTSLSMFSKSEFSVKRRPFDGLHSLKSKYL 199
DB 232 -ELELEAENRKLLTEKDAQIA-----MMQORIDRLALLNEKQAAS-- 271
QY 200 HVGIVYPPAPEKSIIVGTVKVKGKEDSSSTFEVFKRRRAALERYLQRTVKHPTLLQDPDLR 259
DB 272 -----PLEPKE-----LEELRDKNESLTMRLL--HETLKQCCDLK 303
QY 260 QFLESSELPRAVNTQALSGAGILRMVNKAA-----DAVNKMTIKMNSDANFEKQOQ 312
DB 304 --TEKSOMDRKINQLSENGDLSFKLREFASHLQQLQDALNELTEHSHKATQEWLEKQAO 361
QY 313 FE-----NLDQQLRKLH-----VSVEALVCH 333
DB 362 LEKLSAALQDKKCLEBNEILOGLKSQLEHLSQDQNPPEKGEVLGVDVQLQETLKQE 421
QY 334 RKELSANTAFA-----AKSAMLGNSEDH-----TALSRLALSQAEV 370
DB 422 AATLAANNLTQARVEMLETERGQOEAKLLAERGHFEKQQLSSLTLDQSSISNLSQA 481
QY 371 EEKIDQLHQEQAFADFYMFSELLSDYIRLIAVKGVDHRMKCWQKVEDAQITLLKKREA 430
DB 482 KEELEQASQAHGARLTAQVASTLSELTTLNATI-----QQOQDELAGLQKQAO 529
QY 431 EAKMWANKPDKTQQA-----KNEIREWAKVQOGERDFEQISKTIRKEVGR----- 477
DB 530 EKAQLAQTLLQOQEQASQGLRHQVEQLSSLSLKQKEQOLKEVAE--KQEAATRODHAQOLAT 587
QY 478 --FEKERVKDFKTVIIKYLESVQ-----TQOQL 504
DB 588 AAEREASLRERDAALKQLEALEKEKALEILQOQL 624

```

RESULT 6  
 US-08-467-781-4  
 Sequence 4, Application US/08467781  
 Patent No. 5780596  
 GENERAL INFORMATION:  
 APPLICANT: TOUKATLY, GARY  
 APPLICANT: LIDGARD, GRAHAM P  
 TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
 TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
 STREET: 125 HIGH STREET  
 CITY: BOSTON  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110  
 COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467,781  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: PITCHER ESO, EDMUND R  
 REGISTRATION NUMBER: 27,829  
 REFERENCE/DOCKET NUMBER: MTP-013  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 248-7000  
 TELEFAX: (617) 248-7100  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2101 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-467-781-4

Query Match 5.5%; Score 145; DB 1; Length 2101;  
 Best Local Similarity 18.2%; Pred. No. 0.0047;  
 Matches 105; Conservative 104; Mismatches 188; Indels 180; Gaps 20;

```

QY 20 DLEDGDLFTSTVSTLESSPSPPEASLPADISANSNGPKPTEVLLDDDDREDFAEATE 79
DB 136 DHDGGLNLEDLENFLQKAPVPTCSSTPEELSPSHQAK-----RETRFLELQK 186
QY 80 EVSLDSPEREPIISSEPSPAVPTPTTLIAPRIEKSMSAPVIFDRSREIEEANGDI 139
DB 187 VASSSSGNN--FLSGSPASPMGDILOT---POQMRRL-----KKQLADERSNRD- 231
QY 140 FDEIGVSDPEKVGDCGNMAYMAYRVTTKTSLSMFSKSEFSVKRRPFDGLHSLKSKYL 199
DB 232 -ELELEAENRKLLTEKDAQIA-----MMQORIDRLALLNEKQAAS-- 271
QY 200 HVGIVYPPAPEKSIIVGTVKVKGKEDSSSTFEVFKRRRAALERYLQRTVKHPTLLQDPDLR 259
DB 272 -----PLEPKE-----LEELRDKNESLTMRLL--HETLKQCCDLK 303
QY 260 QFLESSELPRAVNTQALSGAGILRMVNKAA-----DAVNKMTIKMNSDANFEKQOQ 312
DB 304 --TEKSOMDRKINQLSENGDLSFKLREFASHLQQLQDALNELTEHSHKATQEWLEKQAO 361
QY 313 FE-----NLDQQLRKLH-----VSVEALVCH 333
DB 362 LEKLSAALQDKKCLEBNEILOGLKSQLEHLSQDQNPPEKGEVLGVDVQLQETLKQE 421
QY 334 RKELSANTAFA-----AKSAMLGNSEDH-----TALSRLALSQAEV 370
DB 422 AATLAANNLTQARVEMLETERGQOEAKLLAERGHFEKQQLSSLTLDQSSISNLSQA 481
QY 371 EEKIDQLHQEQAFADFYMFSELLSDYIRLIAVKGVDHRMKCWQKVEDAQITLLKKREA 430
DB 482 KEELEQASQAHGARLTAQVASTLSELTTLNATI-----QQOQDELAGLQKQAO 529
QY 431 EAKMWANKPDKTQQA-----KNEIREWAKVQOGERDFEQISKTIRKEVGR----- 477
DB 530 EKAQLAQTLLQOQEQASQGLRHQVEQLSSLSLKQKEQOLKEVAE--KQEAATRODHAQOLAT 587
QY 478 --FEKERVKDFKTVIIKYLESVQ-----TQOQL 504
DB 588 AAEREASLRERDAALKQLEALEKEKALEILQOQL 624

```

RESULT 7  
 US-08-195-487-4  
 Sequence 4, Application US/08195487  
 Patent No. 5783403  
 GENERAL INFORMATION:

```

Db 482 KEELEQASQAHGARLPAQVASLTSELTTLNATI-----QQQOQELAGLKQQAk 529
Qy 431 EAKMVAWANKPKIQQA-----KNBIREWEAKVQOGRDFEQISTIRKEVGR-----477
Db 530 EKQAQLAQTQQCEQASQGLRHQVEQLSSLSLKQKEQOLKEVAS--KQATRODHAQOLAT 587
Qy 478 --FEKERVKDFKTVIKYLESLVQ-----TQQQL 504
Db 588 AABEREASLRDAALKQLEALEKEKAAKLEILQQQL 624

RESULT 8
US-08-483-924-4
; Sequence 4, Application US/08483924
; Patent No. 5882876
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,924
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

NAME: PITCHER ESQ, EDMOND R
REGISTRATION NUMBER: 27,899
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-924-4

Query Match 5.5%; Score 145; DB 2; Length 2101;
Best Local Similarity 18.2%; Pred. No. 0.0047;
Matches 105; Conservative 104; Mismatches 188; Indels 180; Gaps 20;

QY 20 DLEGEDLFTVTVLTLESSPPSPASLPADISANSNGPKPTPEVLWDDDDREDLFARATE 79
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 136 DHEDGLNLNDELNFLOKAPYSTCSSTFPBELSPPSHQAQ-----REIRFLELQK 186
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 80 EVSLDSPEREPILSGESPATVPYPTTLIAPRTESKSMSAPVIFDRSREIEEANGDI 139
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 187 VASSSGNN--FLSGSPASPMGDILQT---PQQMRL-----KKQLADERSNRD- 231
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 140 FDIIGVSDPEKVGDMGNMAYRVTTTKTSLSMFSKEFSVKRRFSDFLGLHSLASKYL 199
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 232 -ELELELAENRKLITKEDAQIA-----MMQQRIDRLALLNEKQAAS-- 271
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 200 HGVIVYPAPKPSIVGMTKVKGVEDSSSTFEVEKRRRAALERYLQTVKHPTLLQDPDLR 259
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 272 -----PLEPQE-----LEELRDKNESLFWRL--HEFLKQCODLK 303
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 260 QFLESSELPRAVNTQALSAGAILRMVNKAA-----DAVNKMTKIMNESDAWFEKQQO 312

```

Db 304 --TEKSDMRKINQLSENGDLSFKLREFASHLOQLDALNELTEHSKATQEWLEKQAQ 361  
QY 313 FE-----NLDOQLRLKH-----VSVEALVCH 333  
Db 362 LEKELSAALODKKLEKNEILQGLSLEHLSQLDNPPQEKGEVGLDVLOLETLKAE 421  
QY 334 RKELSANTAAP-----AKSAAMLGNSDEH-----TALSRAISQAEV 370  
Db 422 AATLAANTQOLQARVEMLETERGOQEAELLAERGHFEKQKQSLSLTDLOSSISNLSQA 481  
QY 371 BEKDOLHQAQAFADYMFSELSYIRLIAAVKGVDFHRMKCWKQKQWEDAQITLLKKREA 430  
Db 482 KEELQAQAHGARLTAQVASTSLTTLNATI-----QOQDELGLAGLKQAK 529  
QY 431 EAKMVANKPKIOQA-----KNEIREWAKVQOGRDFEQISKTIRKEVGR 477  
Db 530 EKQAQLAOTLQOQEAQSGRLHQVQLSSSLKQKQKQAEV--KQATRODHAQALAT 587  
QY 478 --FEKERVKDFKTVIKYLESVQ-----TOOOL 504  
Db 588 AAEERASLRERDAALKOLEALEKAKEAKLEILOOOL 624

## RESULT 9

US-09-452-294-1  
; Sequence 1, Application US/09452294  
; Patent No. 6287790  
; GENERAL INFORMATION:  
; APPLICANT: Lelievre, Sophie  
; APPLICANT: Bissell, Mina  
; TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED  
; TITLE OF INVENTION: THERAPY AND DETECTION OF PROLIFERATIVE AND  
; TITLE OF INVENTION: DIFFERENTIATION DISORDERS  
; FILE REFERENCE: IB-1454- Sequence Submittal  
; Patent No. 6287790  
; CURRENT APPLICATION NUMBER: US/09/452,294  
; CURRENT FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: 60/110,420  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2101  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-452-294-1

Query Match 5.5%; Score 145; DB 4; Length 2101;  
Best Local Similarity 18.2%; Pred. No. 0.0047;  
Matches 105; Conservative 104; Mismatches 188; Indels 180; Gaps 20;  
QY 20 DLEDGEDLTSTVSTLESSPSPPEASLPADISANSNGPKTEVTVVLDDEDLFAEATE 79  
Db 136 DHDGGLNLEDNLFQKAPVSTCSSTPEELSPSHQAK-----REIRFLELQK 186  
QY 80 EYSLDPEREPILSSPSPPAVTPVTTLIAPRIEKSMSAPVIFDRSREEEEEEANGDI 139  
Db 187 VASSSGNN--FLSGSPASPMGDILQ-----PQFQNRRL-----KKQADERSNRD- 231  
QY 140 FDEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSVRRFSDFLGLHSLKASKYL 199  
Db 232 -ELELEAENRKLLEKDAQIA-----MMQQRIDRLALLNEKQAA-- 271  
QY 200 HVGIIYPPAPEKSVGTWKVKGEDSSSTFEVKKRALLERYLQRTVTKHPTLLQPDRL 259  
Db 272 -----PLEPKE-----LEELDRKNESLTMRL--HETLKQCCDLK 303  
QY 260 QFLESSELPRAVNTQALSGAILRMVNKAA-----DAVNKMTIKMNSDAWFEKQOQ 312  
Db 304 --TEKSDMRKINQLSENGDLSFKLREFASHLOQLDALNELTEHSKATQEWLEKQAQ 361  
QY 313 FE-----NLDOQLRLKH-----VSVEALVCH 333

Db 362 LEKELSAALODKKLEKNEILQGLSLEHLSQLDNPPQEKGEVGLDVLOLETLKAE 421  
QY 334 RKELSANTAAP-----AKSAAMLGNSDEH-----TALSRAISQAEV 370  
Db 422 AATLAANTQOLQARVEMLETERGOQEAELLAERGHFEKQKQSLSLTDLOSSISNLSQA 481  
QY 371 BEKDOLHQAQAFADYMFSELSYIRLIAAVKGVDFHRMKCWKQKQWEDAQITLLKKREA 430  
Db 482 KEELQAQAHGARLTAQVASTSLTTLNATI-----QOQDELGLAGLKQAK 529  
QY 431 EAKMVANKPKIOQA-----KNEIREWAKVQOGRDFEQISKTIRKEVGR 477  
Db 530 EKQAQLAOTLQOQEAQSGRLHQVQLSSSLKQKQKQAEV--KQATRODHAQALAT 587  
QY 478 --FEKERVKDFKTVIKYLESVQ-----TOOOL 504  
Db 588 AAEERASLRERDAALKOLEALEKAKEAKLEILOOOL 624

## RESULT 10

PCT-US93-06160-4  
; Sequence 4, Application PC/TUS9306160  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA HURWITZ & THIBEAULT  
; STREET: 53 STATE STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/06160  
; FILING DATE: 19930621  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESO, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/248-7000  
; TELEFAX: 617/248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US93-06160-4

Query Match 5.5%; Score 145; DB 5; Length 2101;  
Best Local Similarity 18.2%; Pred. No. 0.0047;  
Matches 105; Conservative 104; Mismatches 188; Indels 180; Gaps 20;  
QY 20 DLEDGEDLTSTVSTLESSPSPPEASLPADISANSNGPKTEVTVVLDDEDLFAEATE 79  
Db 136 DHDGGLNLEDNLFQKAPVSTCSSTPEELSPSHQAK-----REIRFLELQK 186  
QY 80 EYSLDPEREPILSSPSPPAVTPVTTLIAPRIEKSMSAPVIFDRSREEEEEEANGDI 139  
Db 187 VASSSGNN--FLSGSPASPMGDILQ-----PQFQNRRL-----KKQADERSNRD- 231  
QY 140 FDEIGVSDPEKVGDMNAYMAYRVTTKTSLSMFSKSEFSVRRFSDFLGLHSLKASKYL 199

Db 232 -ELELELAENRKLITLTKDAQIA-----MMQQRIDRLALLNEKQAAS-- 271  
QY 200 HVGIVYPPEKSVGKTKVYKVGKEDSSSTFEVFKRAALERYLQRTVKPTLLQDPDLR 259  
Db 272 -----PLEPKE-----LEELRDKNESLTMRL--HETLKQCODLK 303  
QY 260 QFLESSLPRAVNTQALSGAGILRMVNKAA-----DAVNKMTIKMNSDAMFEERKQOQ 312  
Db 304 --TEKSDMRKINQISENGDLSFKLEPFAHSHLQLODQALNELTEHSKATQEWLEKQAO 361  
QY 313 FE-----NLQDLRLKHL-----VSVEALVCH 333  
Db 362 LEKLSAALQDKKLEBKNEILQGLKLSQLEHLSQLQDNPPEKGEVGLQVQLETLKQE 421  
QY 334 RKELSANTAAP-----AKSAAMLGNSDDH-----TALSRLSQAIEV 370  
Db 422 AATLAANNTOLOARVEMLETERGOOEAKLAERGHFEERKQOQSSLTITDQSSISNLSQA 481  
QY 371 EEKIDOLHQBQAEADFYMFSELLSDYIRLIAAVKGVDFHRMKCWKQWEDAQITLLKKREA 430  
Db 482 KELEQASQAHGARLTAQVASTSELTTLNATI-----QQQDELAKLQKQAK 529  
QY 431 EAKMVANKPKIQQA-----KNEIREWAKVQOGERDFEQISKTIRKEVG-- 477  
Db 530 EKAQLAQTLQOQEQASQGLRHQVQQLSSSLKQEQOLKEVAE--KQEAATRODHAQOLAT 587  
QY 478 --PEKERVKDFKVIKIYLSLVQ-----TQOQL 504  
Db 588 AABEREASLRERDAALQKLEALEKAKLEILQOQL 624

## RESULT 11

US-08-938-105-3

; Sequence 3, Application US/08938105

; Patent No. 6353151

; GENERAL INFORMATION:

; APPLICANT: Leinwand, Leslie A.

; APPLICANT: Vikstrom, Karen L.

; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheridan Ross P.C.

; STREET: 1700 Lincoln St., Suite 3500

; CITY: Denver

; STATE: CO

; COUNTRY: U.S.A.

; ZIP: 80203

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/938,105

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Crook, Wanneil M.

; REGISTRATION NUMBER: 31,071

; REFERENCE/DOCKET NUMBER: 3595-4

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 863-9700

; TELEFAX: (303) 863-0223

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1886 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-938-105-3

Query Match 5.58; Score 143.5; DB 4; Length 1886;  
Best Local Similarity 18.78; Pred. No. 0.0054;

Matches 108; Conservative 113; Mismatches 201; Indels 157; Gaps 27;

QY 19 EDLEDGEDLETSVTSVLESPPSPERASLPASISANSNGPKTEVVLDDDDREDFAEAT 78  
Db 1333 EELEEAARKKLAQLQDAEEA-----VEAVNAKSSLEKTKHRLQNEIEDLWYDVE 1382

QY 79 EE-----VSLDSPER--EPILSS-----EPSPAVTPVPTTLIAPRIESKMSAPVI--- 123  
Db 1383 RSNAAAALDKKRNFDKILAEWKQYEE-----QSELESQKREARSLSTELFKLN 1435

QY 124 -----FORSREIEEEANGDIFDIEIGVSD--PERKVG--GMNAYMAYRVTTKSL 170  
Db 1436 AYESLEHLETFKRNKNLQEE-----ISDLTEQLGEGGKNVHELEKI----- 1478

QY 171 SMFSKSEFSVKRRPFDGLGHSKLASKYLHVGIYVPAPEKSIWGTVKVKGEDSSSTE 230  
Db 1479 -----RKOLEVEK-----LELQSAL-----EEAEASLE 1501

QY 231 FVEKR--RAALE--RYLQRTVKHPTLLQDPDLROFLES-----SELPRAVNTQALSGAGIL 282  
Db 1502 HEGKILRAQLEFNQIKAEIERKLAEDMEQAKRNHRLRVDSLQTSLSDAETRSRNEAL 1561

QY 283 RMVNAADAVANKMTIKMNSDAMFEERKQOQFENLDQQLRKLHVSVEALVCHRRKELSANTA 342  
Db 1562 RVKKMEGDLNEMEIQLSQANRIASEAKHLKNAQAHLKDTQLQDLDVAVRANDDLKENIA 1621

QY 343 AFAKSAAMLGNSE-----DHTALSRLS--QLAEVEEKIDOLHQBQAF----- 383  
Db 1622 IVERRNTLL--OAELEELRAVVEQTERSRLAEQELIETSERVQLLHSQNTSLINQKKMD 1680

QY 384 ADFYMFSELLSDYIRLI-----AAVKGVFDRMKC--WOKWED--AQITLLKKREAE 431  
Db 1681 AD---LSQLQTEVEEAQVQECRNEAEKAKAITDAAMMAEELKEQDTSALHRKKNMEQ 1737

QY 432 AKMVANKPKIQQ-----AKNEIREWAKVQOGERDFEQISKTIRKEVG--RFEKERVK 484  
Db 1738 TIKDLQRLDEAEQIALKGKGGKQKLEAVRELENELEAEOKRNABSVKGMKRKRRIK 1797

QY 485 DF-----KTVIKIYLSLVQOQQLIKYWEAFLPEAK 516  
Db 1798 ELTYQTEEDKKNLVR--LQDLVDLQQLKAYKAYKQAOAEAE 1835

## RESULT 12

US-09-104-324B-4

; Sequence 4, Application US/09104324B

; Patent No. 6232460

; GENERAL INFORMATION:

; APPLICANT: T Rec, Ozlem; Sahin, Ugur; Pfreundschuh, Michael

; TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,

; TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fulbright &amp; Jaworski LLP

; STREET: 666 Fifth Avenue

; CITY: New York City

; STATE: New York

; ZIP: 10103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/104,324B

; FILING DATE: 25-June-1998

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/892,702

; FILING DATE: 15-July-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 6232460man D.

REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5491  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 318-3000  
TELEFAX: (212) 752-5958  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 976 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-104-324B-4

Query Match 5.1%; Score 134.5; DB 4; Length 976;  
Best Local Similarity 18.7%; Pred. No. 0.012;  
Matches 103; Conservative 101; Mismatches 200; Indels 147; Gaps 21;

QY 6 EPPPLGDKPTDFDL-----EDGDLFTST-----VSTLESSPSPSPASLPADISAN 55  
DB 24 KQOTLG-GDSTFKFNKCTEDDLEFPFAKTNLSKNGENIDSDPALQKVNFLPVLQOVGN 82  
QY 56 SN-----GPKPTEVLDLDDREDFAEATEE-----VSLDSPEREPTLSSESPAVTP 102  
DB 83 SDCHQVQGLKDSLENSEGLSRVFSKLYKEAKIKKKVSTAEALRQK-----130  
QY 103 VTPTTLIAPRIEKSMSAPVIFDRGREETEEBANDG---IFDIEIGVSDPEKVGDMNAY 159  
DB 131 -----ESKQENRKIIIEAOKAKIQLQFQNEKVSLEKIEQIENKOLIKENAT 179  
QY 160 M-----AYRTTTSLSMFSKSEFSVKKRRSDPLGLHLSKLSKYLHVGIVVPPAPEKSI 215  
DB 180 RHLNLLKTCARSAEKTKKYEEREETROVYMDLNNNT-----EKMITA 224  
QY 216 MTWKVYKGDSDSTSEFEKRRALERYLQRTVKHPTLLQDPLDPLQFLESSELPRAVNTQA 275  
DB 225 FQLRVAENS-----RLEMHFKLDEYKIQHLE-QEYKKEINDKE 265  
QY 276 LSGAGILRMVNAADAVNKNMTKNES-----DAWFEKQOQFENLDQOLRKLH----- 324  
DB 266 KOVSLLIQITTEKENMKDITLLEESRQVNOLEKTKLOSENKQSTKQHHLTKELE 325  
QY 325 ---VSVEALVCHRKELSAATAFAKSAAMLSGNSDHTALSRLASQALAEVEE-KIDOLHGE 380  
DB 326 DKVLSQSVSTOKAL-----EEDLQIATKICQTEKETQMEESNKA 369  
QY 381 QAFADYMFSELLSDYIRLIAAVKGVDFHRMKQWKQEDQAITLLKKRAEAKMMVANKP 440  
DB 370 RAHSF-----VYTEFTVCSLEELLRTQQ-----RLEKNEDQLKILTM--- 410  
QY 441 DKTOQAKNEIREWEAKVQOQGERDFOEISKTI-RKEVGRFEKERVKDFKVIKYLESLVQ 499  
DB 411 -ELQKSSLEEMTKLTNNKEVELEELKKVLGKETLLYEN---KQFE---KIAELAG 462  
QY 500 TQOOLIKYWEA 510  
DB 463 TEQELIGLLQA 473

RESULT 13  
PCT-US93-03077-1  
Sequence 1, Application PC/TUS9303077  
GENERAL INFORMATION:  
APPLICANT: Board of Regents, The University of Texas System  
APPLICANT: Gaynor, Richard B.  
APPLICANT: Wu, Foon Kin  
TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR  
TITLE OF INVENTION: REGULATING GENE EXPRESSION  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA

ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03077  
FILING DATE: 19930331  
CLASSIFICATION:  
PRIOR APPLICATION NUMBER: US/07/862,025  
FILING DATE: April 2, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kammerer, Patricia A.  
REGISTRATION NUMBER: 29,775  
REFERENCE/DOCKET NUMBER: UTFFD270PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-787-1540  
TELEFAX: 713-749-2679  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1093 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US93-03077-1

Query Match 5.0%; Score 131.5; DB 5; Length 1093;  
Best Local Similarity 19.2%; Pred. No. 0.024;  
Matches 118; Conservative 103; Mismatches 234; Indels 159; Gaps 26;

QY 4 BREPPLGD-----GKPTDFEDLEGDLEFT-----STVSTLESSP-----SSPEASLPAD 51  
DB 69 EPQSPPIASPKAITPVRTVYDESENFSAFLSPDVOTIQKSPVVSPPAKSPORPEE 128  
QY 52 ISAN-----SNGPKPTEVVLDDDR-----EDLFAEAT-----EVSILDSR 88  
DB 129 VKSSLHESLHQSRPTETTESQVKDSSLCVSGETLAACTSPKTEGKHEETVKNESDMK 188  
QY 89 EPLSESPSPAVTPPTTLIAPRIEKSMSAPV---IFDRSEETEE-----134  
DB 189 VPTVSLKVESVIVDKTMTESISNTQSLTAEATKDIALEPKREKHEHQSNTPSPVST 248  
QY 135 -ANGIDFDEIGVSDPEKVGDMNAYMAYRVT-----TKTSLSMFSKSEFSVKRRFS 185  
DB 249 FSSGTTSTSDIEVLHDSVISESSASSRQETDTSKSLHMQTSFOLLASACPEYNRLD 308  
QY 186 DF-----LGHSLKSLKYLHVGIVVPP-----A 208  
DB 309 DFQKLTSCSSDAFERIDFSVQSLDSRSVSEINSDDELSGK---GVALVPIVNSST 364  
QY 209 PEKSVGMTKVKVGKEDSSSTFEVRRALERYLQRTVKHPTLLQDPLDPLQFLESSELP 268  
DB 365 PKSKTVESAEGK---SEEVNETLVIPTEEAEMESGRSAT---PYNCEQPD---LVSS--- 414  
QY 269 RAVNTQALSGAGILRMVNAADAVNKNMTKNESDAWFEKQOQFENLDQOLRKLHVSVE 328  
DB 415 ----TPINEGQTVLDKVAEOCEPAESQEPALSEK-----EDVKTVEFLNEKLEKRAQLL 466  
QY 329 ALVCHRKELSAATAFAK-SAAMLGNSDHTALS-----RALSQALAEVEEKIDOLHGEQAF 383  
DB 467 SL---SKKALLAEAFDNLKDEMFRVKESSSISLKDDEFTQRIAEAKKVLQACKERDA 523  
QY 384 ADYMFSELLSDYIRLIAAVKGVDFHRMKQWQ-----KWEDAQIT-LLKKREAEAKMVA 437  
DB 524 AK-----KEIKNIKEELATRLNSSETADLLKKEDEQIRGLMEEGEKLSKQLH 571  
QY 438 K-----PDKIOAKNEIREWEAKVQOQGERDFOEISKTI---RKEVGRFEKERVKDFKT 489  
DB 572 NSNIKKLRKAKORENNVAKLNKKVELEELQHLKQVLDGKEEVEKQHREN-----624





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:58:57 ; Search time 12.4234 Seconds  
(without alignments)

1761.695 Million cell updates/sec

Title: US-09-763-902b-3

Perfect score: 2620

Sequence: 1 MAAREPPPLGDKPTDFED.....TQQLIKYWEAFLPEAKAIA 519

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pap.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pap.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pap.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pap.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pap.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pap.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pap.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pap.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pap.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pap.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pap.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 327   | 12.5        | 68     | 10 | US-09-864-761-42106 |
| 2          | 323.5 | 12.3        | 422    | 9  | US-10-114-893-71    |
| 3          | 306   | 11.7        | 61     | 9  | US-09-764-868-993   |
| 4          | 248   | 9.5         | 414    | 10 | US-09-925-302-706   |
| 5          | 248   | 9.5         | 414    | 10 | US-09-927-738-8     |
| 6          | 165   | 6.3         | 278    | 9  | US-09-764-868-1051  |
| 7          | 160   | 6.1         | 198    | 10 | US-09-927-738-7     |
| 8          | 139   | 5.3         | 909    | 9  | US-09-925-299-988   |
| 9          | 139   | 5.3         | 909    | 10 | US-09-925-299-988   |
| 10         | 138.5 | 5.3         | 1805   | 9  | US-09-820-843A-73   |
| 11         | 136   | 5.2         | 241    | 10 | US-09-925-302-619   |
| 12         | 135   | 5.2         | 294    | 10 | US-09-927-738-9     |
| 13         | 134   | 5.1         | 856    | 9  | US-09-815-242-11489 |
| 14         | 134   | 5.1         | 1013   | 9  | US-10-076-622-553   |
| 15         | 134   | 5.1         | 1013   | 12 | US-10-007-805-553   |
| 16         | 132.5 | 5.1         | 1884   | 10 | US-09-785-770A-17   |
| 17         | 132.5 | 5.1         | 1907   | 10 | US-09-785-770A-16   |
| 18         | 132   | 5.0         | 1002   | 9  | US-10-076-622-475   |
| 19         | 132   | 5.0         | 1002   | 10 | US-09-604-287A-475  |

20 132 5.0 1002 12 US-10-007-805-475  
21 132 5.0 1095 9 US-10-076-622-493  
22 132 5.0 1095 12 US-10-007-805-493  
23 130.5 5.0 396 10 US-09-764-898-275  
24 130.5 5.0 1501 10 US-09-924-154-17  
25 130 5.0 1616 9 US-09-820-843A-16  
26 129.5 4.9 1341 9 US-10-076-622-565  
27 129.5 4.9 1341 12 US-10-007-805-565  
28 129.5 4.9 1349 9 US-10-076-622-573  
29 129.5 4.9 1349 12 US-10-007-805-573  
30 129 4.9 1945 9 US-09-927-597-2  
31 129 4.9 1979 9 US-09-927-597-4  
32 128.5 4.9 1164 10 US-09-950-046A-2  
33 128 4.9 2462 9 US-09-819-104A-5  
34 127.5 4.9 1786 9 US-09-742-096-3  
35 126.5 4.8 619 9 US-09-882-774-1  
36 125.5 4.8 630 9 US-09-742-096-5  
37 125.5 4.8 888 9 US-09-893-519A-73  
38 124 4.7 496 10 US-09-784-249-2  
39 124 4.7 758 9 US-09-925-299-859  
40 124 4.7 758 10 US-09-925-299-859  
41 124 4.7 1242 9 US-09-925-299-911  
42 124 4.7 1242 10 US-09-925-299-911  
43 123.5 4.7 1043 10 US-09-946-805-4  
44 123 4.7 276 9 US-09-764-868-811  
45 123 4.7 854 10 US-09-815-242-5192

#### ALIGNMENTS

#### RESULT 1

US-09-864-761-42106  
; Sequence 42106, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 42106

LENGTH: 68

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC008458.3

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2

OTHER INFORMATION: EST HUMAN HIT: BE280479.1, EVALUE 2.00e-24

OTHER INFORMATION: SWISSPROT HIT: Q60396, EVALUE 5.10e-01

US-09-864-761-42106

Query Match 12.5%; Score 327; DB 10; Length 68;

Best Local Similarity 100.0%; Pred. No. 1.4e-16;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 LSANTAFAKSAAMLGNSDHTALSRLSOLAEEVEEKIDQLHQBOAFADYMFSELSY 396

Db 1 LSANTAFAKSAAMLGNSDHTALSRLSOLAEEVEEKIDQLHQBOAFADYMFSELSY 60

QY 397 IRLIAAVK 404

Db 61 IRLIAAVK 68

RESULT 2

US-10-114-893-71

Sequence 71, Application US/10114893

Publication No. US20020193567A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M.

APPLICANT: LaVallie, Edward R.

APPLICANT: Collins-Racie, Lisa A.

APPLICANT: Evans, Cheryl

APPLICANT: Merberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Bowman, Michael R.

APPLICANT: Spaulding, Vikki

APPLICANT: Kelleher, Kerry S.

APPLICANT: Genetics Institute, Inc.

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

FILE REFERENCE: GI 6000-10A

CURRENT APPLICATION NUMBER: US/10/114,893

CURRENT FILING DATE: 2002-04-02

EARLIER APPLICATION NUMBER: 09/413,232

EARLIER FILING DATE: 1999-10-06

NUMBER OF SEQ ID NOS: 321

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 71

LENGTH: 422

TYPE: PRT

ORGANISM: Homo sapiens

US-10-114-893-71

Query Match

12.3%; Score 323.5; DB 9; Length 422;

Best Local Similarity 25.8%; Pred. No. 2.9e-15;  
Matches 117; Conservative 83; Mismatches 178; Indels 75; Gaps 11;

QY 88 REPILSSPSPAVTPPTTLIAPRIESKMSAPVIFDRSREEIEEANGDIFDIETGVS 147

Db 9 RRPILLESRVAMAAPVLPQLOOEEDRSKLSRSVDLVNPDLSQ-----IDIPDALS 58

QY 148 DPKVGDGMAYMAYRVTTTSLSMFSKSESVKRRSDFLGLHSLKSLKYLHVGYIVPP 207

Db 59 ERDKV-----KFTVHTTKTTLPTFQSPSEFSVTRQHEDFVWLHDTLIETDVAAGLIIPP 110

QY 208 APEKSIVG-----MTKVKGKEDSSSTFEVEKRR-----AALERYLQRTVK 248

Db 111 APTKPDGDPREKMGKEGEGSMTEKEEFQOMQOELEAEYLAVFKTKTVSSHEVFLQLSS 170

QY 249 HPTLLQDPDLRQFLESSE--LPRAVNTQALSGAGILRWNVKAADAVNMKTKMNESDAWF 306

Db 171 HPVLKSDRNFHFVLEYDQDLSVRKNTKEMFG-GFFSVVKSADDEVLTGVK--EVDDEF 227

QY 307 BEKQQPENLDQQLRKHVSVEALVCHRKELS---ANTAFAKSAAMLGNSDHTALSRA 363

Db 228 EOEKNFLINYNRKIDSCVKADKTRSHKNVADDYIHTAACLSLAL-----EPTVIKKY 283

QY 364 LSOLAEEVEEKIDQLHQBOAFADYMFSELSYIRLIAAVKGVFDHMKCKWKWEDAOIT 423

Db 284 LKVAELFEKLRKVEGRVSSDEDLKTELLRYMLNIEAAKDLRYRTRKALIDYENSNA 343

QY 424 LKKREAEAKMVMVANKPKIQQAQKNEIREWEAKVQOQERDFEQISKIRKEVGRFEKERV 483

Db 344 L-----DKARLSKQVKLAFAHQEQCCQKFEQLSESKEELINFRKRV 387

QY 484 KDKRTVIKKYLE-----SLVQFQQOLIK 506

Db 388 AAFKRNLIENSELEIKHARNNVSLQSCIDLFK 420

RESULT 3

US-09-764-868-993

Sequence 993, Application US/09764868

Patent No. US20020168711A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT232

CURRENT APPLICATION NUMBER: US/09/764,868

CURRENT FILING DATE: 2001-01-17

Prior application data removed - refer to PALM or file wrapper

NUMBER OF SEQ ID NOS: 1510

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 993

LENGTH: 61

TYPE: PRT

ORGANISM: Homo sapiens

US-09-764-868-993

Query Match

Best Local Similarity 100.0%; Pred. No. 3.5e-15;

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 KRRESDFLGLHSLKSLKYLHVGYIVPPAPEKSIVGTMKVKGKEDSSSTFEVEKRAALE 240

Db 1 KRRESDFLGLHSLKSLKYLHVGYIVPPAPEKSIVGTMKVKGKEDSSSTFEVEKRAALE 60

QY 241 R 241

Db 61 R 61

RESULT 4

US-09-925-302-706

Sequence 706, Application US/09925302

Patent No. US20020044941A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA104  
CURRENT APPLICATION NUMBER: US/09/925,302  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05918  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 896  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 706  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-302-706

Query Match 9.5%; Score 248; DB 10; Length 414;  
Best Local Similarity 22.9%; Pred. No. 5.7e-10;  
Matches 92; Conservative 74; Mismatches 170; Indels 66; Gaps 9;  
  
QY 125 DRSREIEEANGDI---FDIEIGVSDPEKVGDMNAYMAYRTTKTSLSMFSKSEFSVK 181  
DB 24 DRGLKAINVDLQSDAALQVDISDALSERDKV-----KFTVHTKSSLPNFKQNEFSV 75  
  
QY 182 RRFSDFLGLHSLASKYLHVGIYVPAPEKSIIVGTMKVKGK-----EDSSSTFEVEKRA 237  
DB 76 RQHEEFILWHDSEFVENEYAGYIIPAPPRPDFDASREKLQKLGEGETMKEEFTKMKQ 135  
  
QY 238 ALER-----YLQRTVKHPTLLQDPDLRQFLESSELPRAVNTQALSGAG-- 280  
DB 136 ELEAEYLAIFKTKTVAMHEVFLCRVAHPILRDLNPHVFEYN-----QDLSVRGKN 187  
  
QY 281 -----ILRMVNAADAVNKMKTIKMNSDAMFEKQOQFENLDQOLKRLHVSVEALVCH 333  
DB 188 KKEKLEDFEFKNMVKSDGVIVSGVK--DVDDDFEHEHTEFLLEVHNRVKDASAKSDRMTRS 245  
  
QY 334 RRELSANTAAPAKSAAMLGNSDHTALSRALSQLAERVEKIDQLHQQEAFADFYMESELL 393  
DB 246 HKSAADDYNRIGSSLYALG--TODSTDICFFLKVSELFDKTRKIEARVSADEDLKLSDLL 304  
  
QY 394 SDYIRLIAAVKGVDFHRMCKWQKWEAQITLLKKREAEAKMMVANKPKDQIOQAKNIREW 453  
DB 305 KYLRSSQAKDLLYRSLVDYENANKAL-----DKRAKNKDVLOA 348  
  
QY 454 EAKVQOGERDFEQISKTRKEVGRFERKERVKDKFTVIIRYLE 495  
DB 349 ETSQQLCCCKFEKISSAKQELIDFKTRRVAAPRKNLVELAE 390

RESULT 5  
US-09-927-738-8  
Sequence 8, Application US/09927738  
Patent No. US20020076799A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Tongwen  
TITLE OF INVENTION: Compositions and Methods of modulating TGF-B Signaling  
FILE REFERENCE: 17633/1082  
CURRENT APPLICATION NUMBER: US/09/927,738  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: US 60/119786  
PRIOR FILING DATE: 1999-02-11  
PRIOR APPLICATION NUMBER: PCT/US00/03561  
PRIOR FILING DATE: 2000-02-11  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 8  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Clone SL12 -2 protein Sequence  
US-09-927-738-8

Query Match 9.5%; Score 248; DB 10; Length 414;  
Best Local Similarity 22.9%; Pred. No. 5.7e-10;  
Matches 92; Conservative 74; Mismatches 170; Indels 66; Gaps 9;  
  
QY 125 DRSREIEEANGDI---FDIEIGVSDPEKVGDMNAYMAYRTTKTSLSMFSKSEFSVK 181  
DB 24 DRGLKAINVDLQSDAALQVDISDALSERDKV-----KFTVHTKSSLPNFKQNEFSV 75  
  
QY 182 RRFSDFLGLHSLASKYLHVGIYVPAPEKSIIVGTMKVKGK-----EDSSSTFEVEKRA 237  
DB 76 RQHEEFILWHDSEFVENEYAGYIIPAPPRPDFDASREKLQKLGEGETMKEEFTKMKQ 135  
  
QY 238 ALER-----YLQRTVKHPTLLQDPDLRQFLESSELPRAVNTQALSGAG-- 280  
DB 136 ELEAEYLAIFKTKTVAMHEVFLCRVAHPILRDLNPHVFEYN-----QDLSVRGKN 187  
  
QY 281 -----ILRMVNAADAVNKMKTIKMNSDAMFEKQOQFENLDQOLKRLHVSVEALVCH 333  
DB 188 KKEKLEDFEFKNMVKSDGVIVSGVK--DVDDDFEHEHTEFLLEVHNRVKDASAKSDRMTRS 245  
  
QY 334 RRELSANTAAPAKSAAMLGNSDHTALSRALSQLAERVEKIDQLHQQEAFADFYMESELL 393  
DB 246 HKSAADDYNRIGSSLYALG--TODSTDICFFLKVSELFDKTRKIEARVSADEDLKLSDLL 304  
  
QY 394 SDYIRLIAAVKGVDFHRMCKWQKWEAQITLLKKREAEAKMMVANKPKDQIOQAKNIREW 453  
DB 305 KYLRSSQAKDLLYRSLVDYENANKAL-----DKRAKNKDVLOA 348  
  
QY 454 EAKVQOGERDFEQISKTRKEVGRFERKERVKDKFTVIIRYLE 495  
DB 349 ETSQQLCCCKFEKISSAKQELIDFKTRRVAAPRKNLVELAE 390

RESULT 6  
US-09-764-868-1051  
Sequence 1051, Application US/09764868  
Patent No. US20020168711A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT232  
CURRENT APPLICATION NUMBER: US/09/764,868  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1510  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1051  
LENGTH: 278  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (6)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (17)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (250)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-1051

Query Match 6.3%; Score 165; DB 9; Length 278;  
Best Local Similarity 21.9%; Pred. No. 0.00022;  
Matches 70; Conservative 55; Mismatches 111; Indels 84; Gaps 15;  
  
QY 33 STLESSPSPASLPADISANSNGPKPEVVLDDDDDLFAEATEEVSLSPEPEPIL 92  
DB 10 SGIPGSTXANDP--LPAAAVGAAAE-----AEADEEA-----DPPA 43

QY 93 SSEPSAVTPTPTTLI-----APRIESKMSAPVIFDRSRSEIEEANGDIFDIEIGVSD 148  
Db 44 SLPPTQA--IEPQAIQOVPAFSPQMPGNPLLSHTLQELLAR---DTVOVEL---I 95  
QY 149 PKVGDGMAYMAYVTTKTSLSMSEKSEFSEKRRFSDFLGHLKSLKYLHVGIVPPA 208  
Db 96 PEKGLFLK-HVEYEVSSQ-----REFSSVRYRNDVVFQEMLLHKFPY--RWPAL 145  
QY 209 PEKSTVGMTKVKVGKEDSSSTFEVKKRAALERYLQRTVKHPTLLQDPLRQFLESSLP 268  
Db 146 PPKRMUG-----ADREFTEARRALKRFVNLVARHPFSESDVVLKFL----- 188  
QY 269 RAVNTQALSGAGILRMVKNADAVN-----KMTKMN-----ESDAWFEKQQOFENLD 317  
Db 189 -----SPSGSDVQNKLESACQGVGDEFNCKLATRAKDFLPADIAQOFAISRELIRNY 242  
QY 318 QOLRKLHVSVEALVC--HRK 335  
Db 243 NSFHLRRAERIACAGHRQ 262

RESULT 7  
US-09-927-738-7  
; Sequence 7, Application US/09927738  
; Patent No. US20020076799A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongwen  
; TITLE OF INVENTION: Compositions and Methods of modulating TGF-B Signaling  
; FILE REFERENCE: 17633/1082  
; CURRENT APPLICATION NUMBER: US/09/927,738  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/119786  
; PRIOR FILING DATE: 1999-02-11  
; PRIOR APPLICATION NUMBER: PCT/US00/03561  
; PRIOR FILING DATE: 2000-02-11  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Clone S1 + 12 Protein  
US-09-927-738-7

Query Match 6.1%; Score 160; DB 10; Length 198;  
Best Local Similarity 28.7%; Pred. No. 0.00031;  
Matches 47; Conservative 25; Mismatches 62; Indels 30; Gaps 4;  
QY 125 DRSREIEEANGDI---FDIEIGVSDPEKVGDMNAYMAYVTTKTSLSMSEKSEFSEV 181  
Db 22 DRGLKAINVDLSAALQVDISALSERDKV-----KFTVHTKSLPNFKQNEFSV 73  
QY 182 RRFSDFLGLHSLKSLKYLHVGIVPPAPEKSTVGMTKVKVGK-----EDSSSEFEVKKRA 237  
Db 74 ROHEFIWLHDSFVENEYAGYIIIPAPRPDPFASREKQLKLGEGESMTKEFTKMKQ 133  
QY 238 ALER-----YLOQRTVKHPTLLQDPLRQFLESSE 266  
Db 134 ELEAEVLAIFKKTVMAMHEVFLCRVAAPHILRRDLNFHFVLEYNQ 177

RESULT 8  
US-09-925-299-988  
; Sequence 988, Application US/09925299  
; Publication No. US20030040617A9  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 988  
; LENGTH: 909  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (32)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (41)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (47)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (48)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (52)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (58)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (62)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (125)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (632)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (851)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-988

Query Match 5.3%; Score 139; DB 9; Length 909;  
Best Local Similarity 20.3%; Pred. No. 0.077;  
Matches 119; Conservative 102; Mismatches 260; Indels 104; Gaps 23;  
QY 3 AEREP--PPLGDKPTDFEDLE----DGEDL--FTSTVSTLESSPS---SPEASLPAED 51  
Db 182 AERSNQCKKKEGAPIQCKKADSVANQGTKEGINTQCKKAEKGSFSEKKAEGSPNQKK 241  
QY 52 ISANSNGPKTEVLLDDDDREDLFAEATEEVSLSDSPEPILS---SESPAVTPTPTTL 108  
Db 242 ADAAANQCKTESASVQGRNTDVAQSPAPKQAPAKKSGSKKKEGPGP---PDADG 298  
QY 109 IAPRIESKMSAPVIFDRSR-----EIEEANGDIFDIEIGVSDPEKVGDMNAYMAYRV 164  
Db 299 YLPYKTLTVSGVMVNEGEAQRLLIEISEKAGILQDT---WHKATQGDVP-ALIKKQL 354  
QY 165 TTKTSLSMSEKSEFSEVKKRRFSDFLGHLKSLKYLHVGIVPPAPEKSTVGMTKVKVGKE 224  
Db 355 EEKEKLLATEQEDAAVAK--SKLRELNKEMA-----EKAKAAGEAKVKKKQ 399  
QY 225 DSSSTFEVKKRAALERYLQRTVKHPTLLQDPLRQFLESSLPRAVNTQAL---SGAGI 281  
Db 400 LVAREQEITAVQARMQASVYREHVKEVQQLQG-KIRTLOE--QLENGPNTQLARQENSI 456  
QY 282 LRMVKNADAVNKTIKM-NESDAWFEKQQOFENLDQOLRKLHVSVEALVCHRRLSAN 340  
Db 457 LR-----DALNQATSOVESQNAELAKLRQELSKVSELVEKSAVRQDEQKALEAK 510  
QY 341 TAAFAKSAAML--GNSDHTALSRLSOLAEEVEEKIDQLHQEQAFADFVWFSELLSDIYR 398



QY 272 NTOALSAGILRMVNKAADAVNMTIKMNSDAWFEKQOQFENLQO----- 318  
Db 735 -----KORIADFHNDLSLAKLNEYELSLQKRLQELQTLQLEANKQHSYQNOAYF 781  
QY 319 --OLRLHVSVEALVCHRLKLSANTAAFAKSAAMLGNSDHTALSALSQLAEEVEEKIDQ 376  
Db 782 EGELDLNREKQAFNLNRKQTMVEYDAIKORL-----SKHQALN---MQQAEFLDRKTHE 833  
QY 377 LHQEQAFADYFSELSYIRLIAVKGVFDRMKCQWQWEDAQITLLKKREAEAKMV 436  
Db 834 LN--NAFLNHDADOKSLQOQ---LATVKET-----QKLIDLESALLEKOREFAENV 880  
QY 437 A-----NKPDKIOAKNEIREWEAKVQOGERDFEQISKTIRKEVGRPEKERVKDFKV 489  
Db 881 AGFKRWSNKSQLOKIYELTKQSEQOQKETELKIAFSDLOKQYQVPELOKQDFEQI 940  
QY 490 IIRY--LESLVOTQOOL 504  
Db 941 EAKQRELDKLAENNOV 957  
RESULT 11  
US-09-925-302-619  
; Sequence 619, Application US/09925302  
; Patent No. US20020044941A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 619  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (214)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (220)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-619  
Query Match 5.2%; Score 136; DB 10; Length 241;  
Best Local Similarity 21.1%; Pred. No. 0.02;  
Matches 49; Conservative 48; Mismatches 85; Indels 50; Gaps 9;  
QY 110 APRIEKSMSAPVIFDRSREIEEANGDIFDIETIGVDPKVGDMNAYMAYVTTKTS 169  
Db 27 SPSNDAAGNPLLSHTIQELLAR---DTVQVEL---IPEKGLFLK-HVEYEVSSQ-- 77  
QY 170 LSWFSKSEFVKRRFSDFLGLHSLKSLKYLHVGYIVPPAPEKSIIVGKTVKVKEDSSST 229  
Db 78 -----RKSSVYRRYNDVFVQEMLLHKFPY--RMVPALPPKRMGL-----ADR 119  
QY 230 EFVEKRAALERYLQRTVKTLLQDPLDRLQFLSESLPRAVNTQALSGAGILRMVNKAA 289  
Db 120 EFIEARRALKRFLVNLVARHPFSEDVVLKFL-----SFGSDVQNKLKESA 167  
QY 290 DAVN-----KMTKKN-----ESDAWFEKQOQFENLQDLKRLHVSVEAL 330  
Db 168 QCVGDFLCKLATRAKDFLPADIAQOAFALSRELIRNIYSFHKLRABRI 219  
RESULT 12

US-09-927-738-9  
; Sequence 9, Application US/09927738  
; Patent No. US2002007679A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongwen  
; TITLE OF INVENTION: Compositions and Methods of modulating TGF-B Signaling  
; FILE REFERENCE: 1763371082  
; CURRENT APPLICATION NUMBER: US/09/927,738  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/119786  
; PRIOR FILING DATE: 1999-02-11  
; PRIOR APPLICATION NUMBER: PCT/US00/03561  
; PRIOR FILING DATE: 2000-02-11  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 9  
; LENGTH: 294  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Clone sl+12 -5 protein  
US-09-927-738-9  
Query Match 5.2%; Score 135; DB 10; Length 294;  
Best Local Similarity 21.0%; Pred. No. 0.031;  
Matches 57; Conservative 55; Mismatches 124; Indels 36; Gaps 5;  
QY 233 EKRRALERYLQRTVKTLLQDPLDRLQFLSESLPRAVNTQALSGAG-----ILR 283  
Db 26 KKTVMHEVFLCRVAHAHPILRRDLNPHVLEYN-----QDLSVRGKNKKEKLEDFEK 77  
QY 284 MVNKAADAVNMTIKMNSDAWFEKQOQFENLQDLKRLHVSVEALVCHRLKLSANTAA 343  
Db 78 NMVKSADGVIVSGVK--DVDDFEHEHTEFLLEVHNRVKDASAKSDRMTRSHKSAADDYNR 135  
QY 344 FAKSAAMLGNSDHTALSALSQLAEEVEEKIDQLHQEQAFADYFSELSYIRLIAV 403  
Db 136 IGSSLYALG-TQDSTDICKFFLKVSELSFKTRIEARVSADEDLKLSDLLKYLRESQAA 194  
QY 404 KGVDFHRMKCQWQWEDAQITLLKKREAEAKMVANPKDRIQOAKNEIREWEAKVQOGERD 463  
Db 195 KDLRYRSRLVDYENANKAL-----DKARAKNDVLQAEISQQLCCQK 238  
QY 464 FQISKTIRKEVGRPEKERVKQFVKTVIIRYLE 495  
Db 239 FEKISESAKQELIDFKTRRVAARFKNLVELAE 270  
RESULT 13  
US-09-815-242-11489  
; Sequence 11489, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Travick, John D.  
; APPLICANT: Cart, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11489  
LENGTH: 856  
TYPE: PRT  
ORGANISM: Helicobacter pylori  
US-09-815-242-11489

Query Match 5.1%; Score 134; DB 10; Length 856;  
Best Local Similarity 18.8%; Pred. No. 0.16; Mismatches 163; Indels 200; Gaps 23;  
Matches 104; Conservative 86;  
QY 50 EDISANSN-----GPKTEVVLDLDDREDLFAEATE-----EVSLSDSPEREILSSEPS 97  
DB 148 QDNDDNSLESLEKFGIDLTQKALENKLPVIGRDEIIRMMQILIRKTKNNPILLGEPG 207  
QY 98 PAVTPV-----TPTLIAPRIESKMSAPVIFDRSR-----EIEEEE 134  
DB 208 VGKTAYVEGLAQRIVNKEVPKTLNKRVRVALDLSLLVAGAKYRGEFEERLKKYIEEVKKS 267  
QY 135 ANGDIPIET-----GVSDPEKVGGMNAYMAYRYTTKLSLMSKSEFSVRRFSDFL 188  
DB 268 ANVILFIDEIHTVAGASE-----GGMDAANI----- 295  
QY 189 GLHSLKSLRYLHGYIVPAPPEKSIYVGMKVKYVKEDSSSTEFEVKRRALERYLQRTVK 248  
DB 296 -LKPALARGELHT-----IGATLK-----EYKYEKDMALQRF----- 330  
QY 249 HPFLLOPDLROFL-----ESELPR--AVNTQALSGAILR-----MVNKAADA 291  
DB 331 QPILLNEPSINEALQILRLKLETHTHHNITINDSALIASAKLSRYITDRFLPDRAIDL 390  
QY 292 VNK--WTIKNEDANFEKQOQFENDLQOLRLHVSVEALVCHRLSANTAAFAKSA 349  
DB 391 IDEGAQLKMQ-----MESEPAKLSVKRSIQRLMEKQALEMENKE----- 432  
QY 350 MLGNSDHTALSRLSOLAEVEKIDQHQEAFADYMFSELLSDYIRLIAAVKGVFDH 409  
DB 433 ----SNHKKMQBILKSLDKKEKIQLE-----AOFENEKEVFKLSRLKMEGGL--- 479  
QY 410 RMKCQKWEDAQITLLKKRAE-----AKMNVANKPKIQQAINEIREWEAKVQ 458  
DB 480 -----KKEAERFKRNGDYQQAIEIYSKIPEKEKKELOHKWETMQQ 522  
QY 459 OG-----ERDFEQISK-----TIRKEVGRFEKERVKDFKTVIILKYLESLVQOQLI 505  
DB 523 NGALLONALTENNIAETIVSOWTHIPQKML-OSEKNRVNLNIES-----ELQKRVV 571  
QY 506 KYWEAFLEAKAI 518  
DB 572 GQEKALAKIAKAI 584

RESULT 14  
US-10-076-622-553  
Sequence 553, Application US/10076622  
Publication No. US20030023036A1  
GENERAL INFORMATION:  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Sleath, Paul R.  
APPLICANT: Persing, David H.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.470G11  
CURRENT APPLICATION NUMBER: US/10/076,622  
CURRENT FILING DATE: 2002-02-13

NUMBER OF SEQ ID NOS: 627  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 553  
LENGTH: 1013  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-076-622-553

Query Match 5.1%; Score 134; DB 9; Length 1013;  
Best Local Similarity 20.1%; Pred. No. 0.2; Mismatches 123; Conservative 102; Indels 182; Gaps 30;  
QY 8 PPLGDKGPTDFEDLEDGEDLFTSTVS-----TLESSPSPSEFASL 47  
DB 169 PP--ESKQDYENSWSDESICETVSQKDVCLPKATHQKIDKINGKLESPNK----- 220  
QY 48 PADDISANNGPK---PTEVVLDLDDREDLFAEATEEVSLSDSPEREILSSEPSAPVTPVT 104  
DB 221 --DGLLKATCGMKVSIPTKALELKMOTFKAEPPGPGPSAFEPATE-MOKSVNPAKLELN 277  
QY 105 PTLIAPRI---ESKS-----MSAPVIFDRSREIE--EANGDIFDIE 143  
DB 278 EQTLRADEILPSESKQDYENSWSDESICETVSQKDVCLPKAAHQEKIDKINGKLE---- 333  
QY 144 IGVSDEPKVGDGM---NAYMAYRYTTKLSLMSKSEFSYK-----RRFSDFLGLHSLA 195  
DB 334 --EGSPGK--DGLKANCGMKVSIPTK-ALELMDMOTFKAEPPGKPSAFEPATEMOKSV 388  
QY 196 SKYLHGYIVPAPPEKSIYVGMKVKYVKEDS-----SSTFEVEKR-----RAALERYLQRT 246  
DB 389 NKALELKNQTLRADEILPSESKQDYENSWSDESICETVSQKDVCLPKAAHQEKIDKI 448  
QY 247 VKHPTLLOPDLROFLES--SELPRAVNTQALSGAILRM----- 284  
DB 449 --NGKLEESPDNDGFLKSPCRMVSIPTKALE---LMDMOTFKAEPPGKPSAFEPATEMQ 503  
QY 285 ---VNKAADAVNKMRTIKMNESDAWF---EEKQOQFENDLQOLRLHVSVEALVC-----H 333  
DB 504 KSVNPAKLELKNQTLR---ADQMFPSESKQKNVENSWSDESICETVSQKDVCLPKATH 560  
QY 334 RKLSANTAAFAKSAAMLGNSDHTALSRLSOLAEVE--EKTIDQLHQBOAFADYMFSE 391  
DB 561 QKEMD-----KISGKLESDTSLSKILDTVHSCERARELQKDHCEORTG----- 603  
QY 392 LLSDYIRLIAAVKGVFDHMKCKQKWEADAQITLLKKREAEAKMNVANKPKIQQAINEIR 451  
DB 604 -----KMEQMK---KFCVLKRLKSEAKS-----QLENQKV 634  
QY 452 EWEAKV-----QOGERDFEQISKTIKKEVGRFEKERVKDFKTVIILKYLESLVQO 501  
DB 635 KWQELCSVRLTUNQEEKRRNADILNEKIREELGRIEDQHRKELE--VKQLEQALRIQ 692  
QY 502 QOLIKYWEAFL 512  
DB 693 DIEKSVESNL 703

RESULT 15  
US-10-007-805-553  
Sequence 553, Application US/10007805  
Patent No. US20020150581A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Yugu  
APPLICANT: Dillon, David C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Xu, Jiangchun  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Hepler, William T.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Durham, Margarita

;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
;; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER

;; FILE REFERENCE: 210121.470C10

;; CURRENT APPLICATION NUMBER: US/10/007.805

;; CURRENT FILING DATE: 2001-12-07

;; NUMBER OF SEQ ID NOS: 593

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 553

;; LENGTH: 1013

;; TYPE: PRT

;; ORGANISM: Homo sapiens

US-10-007-805-553

Query Match 5.1%; Score 134; DB 12; Length 1013;

Best Local Similarity 20.1%; Pred.No.0.2; Indels 182; Gaps 30;  
Matches 123; Conservative 102; Mismatches 204;

```
QY 8 PPLGDKPTDFEDGEDGELFTSTVS-----TLESSPSSPEPASL 47
DB 169 PP-ESKQDYEEENWDSSESLCETVSQKDVCLPKATHQKIDKINGKLEESPNK-----220

QY 48 PAEDISANSNGPK---PTEVLDDDDREDFAEATEEVSILDSPEPEILSSEPSAVTPTV 104
DB 221 --DGLLKATCGMKVSIPTKALELKDQTFKAEPGKPSAFEPATE-MOKSVENKALELKN 277

QY 105 PTLIAPRI---ESKS-----MSAPVIFDRSREEIP-EEANGDIEDIE 143
DB 278 EOTLRADEILPSEKQDYEEENWDSSESLCETVSQKDVCLPKAAHQKEIDKINGKKL----333

QY 144 IGVSPKQKVGDM---NAYMAYRVTTKTSLSMFSSEFSVK-----RRFSDFLGLHLSKLA 195
DB 334 --EGSPGK--DGLLKANCGMKVSIPTK-ALELMDQTFKAEPPEKPSAFEPAPAEIMQKSV 388

QY 196 SKYLHVGIVVPAPKPSIVGMTKVKVGKEDS-----SSTEFVEKR-----RAALERYLQRT 246
DB 389 NKALELKNQOTLRADAILPSEKQDYEEENWDSSESLCETVSQKDVCLPKAAHQKEIDKI 448

QY 247 VKHPTLLQPDPLRQFLES-SELPRAVNTQALSGAILRM-----284
DB 449 --NGKLEESPDNDGFLKSPCRMKVSIPTRALE---LMDQTFKAEPPEKPSAFEPAPAEIMQ 503

QY 285 ---VNKAADAVNKMVKIKNESDAWF--BEKQOOFENLDQQLRKLHVSV-EALVC-----H 333
DB 504 KVPNKALELKNQOTLR--ADQMPFSEKQKNVEENWDSSESLRETVSQKDVCPKATH 560

QY 334 RKELSANTAFAKSAAMLGNSEDHTALSALQAEVE--EKIDQLHQQAFADFYMFSE 391
DB 561 QKEMD-----KISGLEDSTSLKILDTVHSCERARELQKDHCEQRTG-----603

QY 392 LLSDYIRLIAVKGVDHRMKQKQWEDAQITLLKKREAEAKMMVANKPKDIOQAKNEIR 451
DB 604 -----KMEQMKK---KFCVLKKKLSAKEIKS-----QLENQKV 634

QY 452 EWEAKV-----QGERDEQISKIRKEVGFERKVKDFKTVIKVLESVQTO 501
DB 635 KWQELCSVRLTLNQBEERKNRADIINERIELGRIEQHRRKELE--VKQOLEQALRIQ 692

QY 502 QQLIKYWEAFL 512
DB 693 DIELKSVESNL 703
```

Search completed: March 4, 2003, 15:07:10  
Job time: 15.4234 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:50:41 ; Search time 61.4251 Seconds

(without alignments)  
1559.740 Million cell updates/sec

\*Title: US-09-763-902b-5

Perfect score: 3638

Sequence: 1 MSVDKAEKCGSLTLWLQTFH.....SRRGLGLASLNLRPTDKH 719

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description         |
|------------|--------|-------------|--------|----|---------------------|
| 1          | 3638   | 100.0       | 719    | 21 | Human protein tran  |
| 2          | 3503.5 | 96.3        | 724    | 22 | Novel human diagno  |
| 3          | 854    | 23.5        | 679    | 22 | Drosophila melanog  |
| 4          | 682.5  | 18.8        | 335    | 22 | Human, diagnostic a |
| 5          | 516    | 14.2        | 105    | 20 | Human 5' EST seque  |
| 6          | 420    | 11.5        | 990    | 22 | Human protein seq   |
| 7          | 391.5  | 10.8        | 931    | 22 | Human protein seq   |
| 8          | 380    | 10.4        | 169    | 22 | Human hook2 protei  |
| 9          | 378    | 10.4        | 1381   | 22 | Drosophila melanog  |
| 10         | 376    | 10.3        | 320    | 22 | Novel human diagno  |

|    |       |     |      |    |          |                    |
|----|-------|-----|------|----|----------|--------------------|
| 11 | 339.5 | 9.3 | 2383 | 23 | ABG65631 | Human breast speci |
| 12 | 339.5 | 9.3 | 2442 | 21 | AAV77575 | Human cytoskeletal |
| 13 | 338.5 | 9.3 | 1203 | 22 | AAV79264 | Human protein sequ |
| 14 | 322   | 8.9 | 1879 | 22 | AAV25750 | Human male enhance |
| 15 | 319   | 8.8 | 1374 | 22 | AAV69070 | Novel human secret |
| 16 | 317   | 8.7 | 2117 | 22 | AAV32040 | Novel human secret |
| 17 | 317   | 8.7 | 2192 | 18 | AAV21732 | LexA/HA/NUA fusio  |
| 18 | 317   | 8.7 | 2272 | 18 | AAV21731 | A human trichohyal |
| 19 | 315   | 8.7 | 1898 | 20 | AAV30795 | Human protein sequ |
| 20 | 315   | 8.7 | 1963 | 22 | AAV79838 | Human protein sequ |
| 21 | 315   | 8.7 | 2115 | 21 | AAV49937 | Novel human secret |
| 22 | 315   | 8.7 | 2207 | 22 | AAV32041 | Novel human secret |
| 23 | 313.5 | 8.6 | 1960 | 22 | AAV78854 | Human protein sequ |
| 24 | 313.5 | 8.6 | 2143 | 22 | ABG01716 | Novel human diagno |
| 25 | 312   | 8.6 | 1154 | 22 | AAV32042 | Novel human secret |
| 26 | 311.5 | 8.6 | 2101 | 15 | AAV71733 | Sequence of the in |
| 27 | 311.5 | 8.6 | 2101 | 22 | AAV71733 | Nuclear mitotic ap |
| 28 | 310   | 8.5 | 1489 | 22 | ABV59948 | Drosophila melanog |
| 29 | 309   | 8.5 | 2246 | 22 | ABG05850 | Novel human diagno |
| 30 | 307   | 8.4 | 2101 | 21 | AAV49936 | Human NUA protein  |
| 31 | 306.5 | 8.4 | 1988 | 22 | AAV40999 | Human polypeptide  |
| 32 | 306.5 | 8.4 | 1988 | 22 | AAV40999 | Human polypeptide  |
| 33 | 304   | 8.4 | 1083 | 22 | ABG20032 | Novel human diagno |
| 34 | 303   | 8.3 | 149  | 22 | ABG07201 | Novel human diagno |
| 35 | 301   | 8.3 | 1388 | 19 | AAV56475 | Protein with Rho p |
| 36 | 301   | 8.3 | 1453 | 22 | AAV39213 | Human polypeptide  |
| 37 | 301   | 8.3 | 1469 | 22 | AAV39214 | Human polypeptide  |
| 38 | 300.5 | 8.3 | 951  | 22 | ABV59033 | Drosophila melanog |
| 39 | 300.5 | 8.3 | 1411 | 17 | AAV02258 | Protein MYH11 diff |
| 40 | 299   | 8.2 | 1857 | 23 | AAU84350 | Protein MYH11 diff |
| 41 | 297.5 | 8.2 | 2067 | 22 | ABV71125 | Drosophila melanog |
| 42 | 297   | 8.2 | 3248 | 17 | AAV99795 | Kinetochore protei |
| 43 | 296.5 | 8.2 | 2779 | 22 | ABV62371 | Drosophila melanog |
| 44 | 296   | 8.1 | 2482 | 16 | AAV72826 | Human mitotin. Ho  |
| 45 | 296   | 8.1 | 2482 | 19 | AAV23996 | Human mitotin amin |

#### ALIGNMENTS

#### RESULT 1

AAV82321

ID AAV82321 standard; Protein; 719 AA.

AC AAV82321;

XX

DT 19-JUN-2000 (first entry)

XX

DE Human protein transport molecule (PTAM) SEQ ID NO:5.

XX Human; protein transport molecule; PTAM; diagnosis; cytostatic;

KW antiarthritic; antiasthmatic; immunosuppressant; antiarteriosclerotic;

KW antiallergic; antidiabetic; antilipemic; antirheumatic; osteopathic;

KW dermatological; antianemic; antipsoriatic; hepatotropic; antigout;

KW antinflammatory; antiHIV; protein transport regulator; cancer;

KW immune disorder; cell proliferative disorder; secretory disorder;

KW urticaria; allergy; abnormal vesicle trafficking; asthma;

KW autoimmune haemolytic anaemia.

XX Homo sapiens.

XX WO200012703-A2.

PN 09-MAR-2000.

XX 26-AUG-1999; 99WO-US19616.

XX 27-AUG-1998; 98US-0098206.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;

PI Gorgone GA, Baughn MR, Patterson C;

XX WPI: 2000-256642/22.  
 DR N-PSDB; AAA08039.  
 XX  
 PT New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis  
 XX  
 PS Claim 1: Page 65-67; 75pp; English.  
 XX  
 CC AAA08035 to AAA08042 encode the human protein transport-associated  
 CC molecules (PTAMs) given in AAY82317 to AAY82324. The PTAMs have  
 CC cytoskeletal, antiarthritic, antiasthmatic, immunosuppressant,  
 CC antitumor, antiallergic, antidiabetic, antilipemic,  
 CC antirheumatic, osteopathic, dermatological, antianemic, antipsoriatic,  
 CC hepatotropic, antigout, antiinflammatory and antihiv activities, and  
 CC regulate protein transport. PTAM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PTAM. PTAM polynucleotides are useful for  
 CC diagnosing conditions associated with PTAM, comprising detecting PTAM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the  
 CC biological sample. Diseases prevented, treated or diagnosed include cell  
 CC proliferative disorders such as cancers, immune disorders, secretory  
 CC disorders and other conditions associated with abnormal vesicle  
 CC trafficking, such as allergies, asthma, urticaria and autoimmune  
 CC haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as  
 CC a targeting or delivery mechanism for bringing pharmaceutical agents  
 CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related  
 CC disorders. PTAM, its catalytic or immunogenic fragments are useful for  
 CC drug screening using libraries of compounds. PTAM polynucleotides are  
 CC useful for generating hybridisation probes useful in mapping the  
 CC naturally occurring genomic sequences.  
 XX  
 SQ Sequence 719 AA;

Query Match 100.0%; Score 3638; DB 21; Length 719;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-253;  
 Matches 719; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVDKAEICGSLTWTQTHVSPSPASPODLSGLAVAVLNQIDPSWNEAWLGISED 60  
 DB 1 MSVDKAEICGSLTWTQTHVSPSPASPODLSGLAVAVLNQIDPSWNEAWLGISED 60  
 QY 61 PGPNNKLVSNLKWLRSLVEYSQDVLAHPVSEELPDVSLGIFSDPAELCKLQVLVG 120  
 DB 61 PGPNNKLVSNLKWLRSLVEYSQDVLAHPVSEELPDVSLGIFSDPAELCKLQVLVG 120  
 QY 121 CAISCEKKQDHIQRIWTLVESQVHVYVMEAIQELMTKDTPTDLSPTYGNFDSQSRYYFL 180  
 DB 121 CAISCEKKQDHIQRIWTLVESQVHVYVMEAIQELMTKDTPTDLSPTYGNFDSQSRYYFL 180  
 QY 181 SEAESEGELOQRCLDLERQLMLLSEKOSLAQENAGLRMRGPRGEGTPTGLTAKKLL 240  
 DB 181 SEAESEGELOQRCLDLERQLMLLSEKOSLAQENAGLRMRGPRGEGTPTGLTAKKLL 240  
 QY 241 LOSOLEQOEENFRLESGRDELRCAELEREVAELQHRNQAULTSLAQEAQAKDMDDEL 300  
 DB 241 LOSOLEQOEENFRLESGRDELRCAELEREVAELQHRNQAULTSLAQEAQAKDMDDEL 300  
 QY 301 ROSSERAGOLEATLTSCRRRLGELRELROVROLEERNAAGHAERTQLEDELRRAGSLRA 360  
 DB 301 ROSSERAGOLEATLTSCRRRLGELRELROVROLEERNAAGHAERTQLEDELRRAGSLRA 360  
 QY 361 QLEAQRQVQELQOQOEAMKAELWFCRNLEEKYESVTEKELLERDSLRANEE 420  
 DB 361 QLEAQRQVQELQOQOEAMKAELWFCRNLEEKYESVTEKELLERDSLRANEE 420  
 QY 421 LRCAQLQPRGLTQADPSLOPTSTPVDNLAAETLPAELRETLRLQLENKRLCQEAADRE 480  
 DB 421 LRCAQLQPRGLTQADPSLOPTSTPVDNLAAETLPAELRETLRLQLENKRLCQEAADRE 480  
 QY 481 ROELQORHLEDANRRHGLETOHRLNQOOLSELRAQVEDLQALQOQGGKTEDAITSILK 540  
 DB 481 ROELQORHLEDANRRHGLETOHRLNQOOLSELRAQVEDLQALQOQGGKTEDAITSILK 540

DB 481 ROELQORHLEDANRRHGLETOHRLNQOOLSELRAQVEDLQALQOQGGKTEDAITSILK 540  
 QY 541 RKEELHLOKLEADLELOKREYIEELPPPTDSSSTARIEELOHNLQKDKADLRAWEERY 600  
 DB 541 RKEELHLOKLEADLELOKREYIEELPPPTDSSSTARIEELOHNLQKDKADLRAWEERY 600  
 QY 601 RRYVDKARMVQTMPEKQRPAGAPPELHSLRTQLRERDVRIRHLEMDFEKSRSQREQEE 660  
 DB 601 RRYVDKARMVQTMPEKQRPAGAPPELHSLRTQLRERDVRIRHLEMDFEKSRSQREQEE 660  
 QY 661 KLLISAWNMGALQORAGEERAPAHASFLAQORLATNSRRGPIGLASLNLRTDKH 719  
 DB 661 KLLISAWNMGALQORAGEERAPAHASFLAQORLATNSRRGPIGLASLNLRTDKH 719  
 RESULT 2  
 ABG10025  
 ID ABG10025 standard; Protein: 724 AA.  
 AC ABG10025;  
 DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #10016.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 QS Homo sapiens.  
 PN WO200175067-A2.  
 XX 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS74212.  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PS Claim 20; SEQ ID No 40384; 103pp; English.  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG0377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 724 AA;

Query Match 96.3%; Score 3503.5; DB 22; Length 724;  
Best Local Similarity 97.4%; Pred. No. 9.3e-244;  
Matches 705; Conservative 0; Mismatches 14; Indels 5; Gaps 5;

QY 1 MSYDKAELCGSLTWTQTHVPSQSPQDSSGLAVAYVNLIDPSWNEALQGISD 60  
DB 1 MSYDKAELCGSLTWTQTHVPSQSPQDSSGLAVAYVNLIDPSWNEALQGISD 60  
QY 61 PGNWKLKVSNLKWLRSVLEYSQDVLAHVPVSEHLPDVSLIGEFSDPAELKQLQVLG 120  
DB 61 PGNWKLKVSNLKWLRSVLEYSQDVLAHVPVSEHLPDVSLIGEFSDPAELKQLQVLG 120  
QY 121 CAISCEKKQDHIQIRIMTLEBSVOHVYMEAIQELMTKDTDPDLSPTYGNFDSRRYYFL 180  
DB 121 CAISCEKKQDHIQIRIMTLEBSVOHVYMEAIQELMTKDTDPDLSPTYGNFDSRRYYFL 180  
QY 181 SEAEAGDELOQRCLOLEROLMILLSEKQSLAENAGLRMRGPGEPTGLTAKKLL 240  
DB 181 SEAEAGDELOQRCLOLEROLMILLSEKQSLAENAGLRMRGPGEPTGLTAKKLL 240  
QY 241 LQSOLEQLOENRLESREDELRCAELREVAELQHRNOALTSIAQAAKQDMDL 300  
DB 241 LQSOLEQLOENRLESREDELRCAELREVAELQHRNOALTSIAQAAKQDMDL 300  
QY 301 RQSSERAGQLEATLTCRRRLGELRELRRQVROLEERNAGHAERTROLELRAGSLRA 360  
DB 301 RQSSERAGQLEATLTCRRRLGELRELRRQVROLEERNAGHAERTROLELRAGSLRA 360  
QY 361 QLEAQRQVQELQGOQOEAMKAEKWLFEENLEEKYESTYKKEKRLLAERDSLRANEE 420  
DB 361 QLEAQRQVQELQGOQOEAMKAEKWLFEENLEEKYESTYKKEKRLLAERDSLRANEE 420  
QY 421 LRCAQLPQGLTQADPSLDTPTSPVDNLAAEILPAELRETLLRLQ-LENKRLCROEAADR 479  
DB 421 LRCAQLPQGLTQADPSLDTPTSPVDNLAAEILPAELRETLLRLCROEAADR 480  
QY 480 ERQEE-LQRHLEDANRHHGLETQHRNLNQO-QLSELRAQVEDLOKALQEOGGKTEDAISI 537  
DB 481 ERQEE-LQRHLEDANRHHGLETQHRNLNQOAYSELRAQVEDLOKALQEOGGKTEDAISI 540  
QY 538 LKRRKEEHQKLKLEHLEQRLQREYIEEPPDSTARRIEELQHNLOKQDADLRAME 597  
DB 541 LKRRKEEHQKLKLEHLEQRLQREYIEEPPDSTARRIEELQHNLOKQDADLRAME 600  
QY 598 ERYRYVDKARMVQMTPEKQRPAAAGAPPELHSLRTOLRERDVIRHLEMDFKRSORE 657  
DB 601 ERYRYVDKARMVQMTPEKQRPAAAGAPPELHSLRTOLRERDVIRHLEMDFKRSORE 660  
QY 658 QEEKLLISAWYNGMALQORAGGERAPAHQS-FLAQORLATN-SRRGPLGLRLASLNLRP 715  
DB 661 QEEKLLISAWYNGMALQORAGGERAPAPCPSPHFLAQORAGNQILARGLRLASLNLRP 720  
QY 716 TDKH 719  
DB 721 TDKH 724

RESULT 3  
ABB71989  
ID ABB71989 standard; Protein: 679 AA.  
XX  
AC ABB71989;  
XX

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 42759.

DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE ) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

N-PSDB; ABL16092.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

Disclosure; SEQ ID NO 42759; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 679 AA;

Query Match 23.5%; Score 854; DB 22; Length 679;

Best Local Similarity 31.7%; Pred. No. 5.4e-53;

Matches 227; Conservative 156; Mismatches 261; Indels 72; Gaps 17;

QY 1 MSYDKAELCGSLTWTQTHVPSQSPQDSSGLAVAYVNLIDPSWNEALQGISD 60  
DB 1 MSYDKAELCGSLTWTQTHVPSQSPQDSSGLAVAYVNLIDPSWNEALQGISD 60  
QY 61 PGNWKLKVSNLKWLRSVLEYSQDVLAHVPVSEHLPDVSLIGEFSDPAELKQLQVL 119  
DB 61 PGNWKLKVSNLKWLRSVLEYSQDVLAHVPVSEHLPDVSLIGEFSDPAELKQLQVL 120  
QY 120 CAISCEKKQDHIQIRIMTLEBSVOHVYMEAIQELMTKDTDPDLSPTYGNFDSRRYY 177  
DB 121 CAISCEKKQDHIQIRIMTLEBSVOHVYMEAIQELMTKDTDPDLSPTYGNFDSRRYY 177  
QY 178 YFLSEEA--REGDELOQRCLOLEROLMILLSEKQSLAENAGLRMRGPGEPTGLTAKK 230  
DB 178 YFLSEEA--REGDELOQRCLOLEROLMILLSEKQSLAENAGLRMRGPGEPTGLTAKK 237  
QY 231 ----PGLTAK-KLLLSQLEQLOENRLESREDELRCAELREVAELQHRNOALTSIA 284  
DB 231 ----PGLTAK-KLLLSQLEQLOENRLESREDELRCAELREVAELQHRNOALTSIA 297  
QY 285 SLAQEAQALKDMDLQRCSSERAGQLEATLTCRRRLGELRELRRQVROLEERNAGHAER 344  
DB 285 SLAQEAQALKDMDLQRCSSERAGQLEATLTCRRRLGELRELRRQVROLEERNAGHAER 357  
QY 345 TRQLEDELRLAGSLRAQLEAQRQVQELQGOQOEAMKAEKWLFEENLEEKYESTYKKE 404  
DB 345 TRQLEDELRLAGSLRAQLEAQRQVQELQGOQOEAMKAEKWLFEENLEEKYESTYKKE 417  
QY 405 ERLLAERDSLRANEEELRCALQOPRGLTQADPSLDTPTSPVDNLAAEILPAELRETLLRL 464  
DB 405 ERLLAERDSLRANEEELRCALQOPRGLTQADPSLDTPTSPVDNLAAEILPAELRETLLRL 464



RESULT 5  
AA12249  
ID AAY12249 standard; Protein; 105 AA.

XX AC AAY12249;

XX DT 18-JUN-1999 (first entry)

XX DE Human 5' EST secreted protein SEQ ID NO: 562.

XX KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
XX KW forensic; gene therapy; chromosome mapping; signal peptide;  
XX KW upstream regulatory sequence; cytokine activity; cell proliferation;  
XX KW differentiation; haematopoiesis regulation; tissue growth regulation;  
XX KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
XX KW thrombolytic; anti-inflammatory; tumour inhibition.

XX OS Homo sapiens.

XX PN WO9906554-A2.

XX PD 11-FEB-1999.

XX PF 31-JUL-1998; 98WO-IB01238.

XX PR 01-AUG-1997; 97US-0905134.

XX PA (GEST) GENSET.

XX PI Duclert A, Dumas Milne Edwards J, Lacroix B;

XX DR WPI; 1999-153784/13.

XX N-PSDB; AAX41082.

XX PT New nucleic acids encoding human secreted proteins - obtained from  
XX PT cDNA libraries prepared from kidney, fetal kidney, dystrophic  
XX PT muscle, muscle and heart tissue

XX PS Claim 34; Page 615-616; 622pp; English.

XX CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for  
XX CC human secreted proteins, and encode the proteins given in AAY01602 and  
XX CC AAY11994 to AAY12260, respectively. The proteins given represent the  
XX CC signal peptide and an N-terminal fragment of a secreted protein. The  
XX CC nucleic acid sequences can be used for producing secreted human gene  
XX CC products. They can also be used to develop products for diagnosis and  
XX CC therapy. The proteins obtained may have cytokine activity, cell  
XX CC proliferation/differentiation activity, haematopoiesis regulating  
XX CC activity, tissue growth regulating activity, reproductive hormone  
XX CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
XX CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
XX CC activity, tumour inhibition activity or other activities. The products  
XX CC can be used in forensic, gene therapy and chromosome mapping procedures.  
XX CC The sequences can also be used for obtaining corresponding promoter  
XX CC sequences. The nucleic acids encoding the signal peptide can be used  
XX CC for directing extracellular secretion of a polypeptide or the insertion  
XX CC of a polypeptide into a membrane, or importing a polypeptide into  
XX CC a cell.

XX SQ Sequence 105 AA;

Query Match 14.2%; Score 516; DB 20; Length 105;  
Best Local Similarity 100.0%; Pred. No. 1e-29;  
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 MYLRSVEVSQDVLHVPVEEHLDPVSLGSDPAELGKLLQLVLGCAISCEKQDHIQ 133

DB 1 MYLRSVEVSQDVLHVPVEEHLDPVSLGSDPAELGKLLQLVLGCAISCEKQDHIQ 60

QY 134 RIMTLESVOHVYMEAIQELMTKDTDPDSLSPETYGNFDSQR 175

DB 61 RIMTLESVOHVYMEAIQELMTKDTDPDSLSPETYGNFDSQR 102

RESULT 6  
AAM78520  
ID AAM78520 standard; Protein; 990 AA.

XX AC AAM78520;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 1182.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX DR WPI; 2001-476283/51.

XX N-PSDB; AAK51653.

XX PT Nucleic acids encoding polypeptides with cytokine-like activities,  
XX PT useful in diagnosis and gene therapy -

XX PS Claim 20; Page 3425-3427; 6221pp; English.

XX CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
XX CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to  
XX CC cytokine, cell proliferation or cell differentiation or which may induce  
XX CC production of other cytokines in other cell populations. The  
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX CC peptide therapy. The polypeptides have various cytokine-like activities,  
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating  
XX CC activity, tissue growth factor activity, immunomodulatory activity and  
XX CC activator/inhibitor activity and may be useful in the diagnosis and/or  
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX CC inflammation.

XX CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
XX CC (AAM80020) are omitted as the relevant pages from the sequence listing  
XX CC were missing at the time of publication.

XX SQ Sequence 990 AA;

Query Match 11.5%; Score 420; DB 22; Length 990;

Best Local Similarity 23.5%; Pred. No. 1.6e-21;

Matches 180; Conservative 156; Mismatches 270; Indels 160; Gaps 29;

QY 12 LITLWLTFFVPSPCASPDQ-----LSSGLAVAVYLNQIDPWFNEAWLOGISDGP 63

DB 5 LYTWTFTF---GFLAAGNTLDEYVALVDGVFLNQVLMQLINPKLESQRNKKVNDAA--59

Qy 64 NWKLVSNLKVLSVLSVSDVLAHPVSEHLPDVSIGLGE--FSDPA--ELGKLLQVLV 119  
Db 60 --SRRMHLNLSLVROIKFYQETLQQLIMMS-LPNVLIIIGKPNFSEQTEEVKKLLLL 116  
Qy 120 GCAISCEKQDHQRIIMTLEESVQHVMEAIQELMTKDTDPDSLSPETVGN--FDSQSR 176  
Db 117 GCAVQCQKKEEFIERIOGLDFTDRAAANAHIQEV-----THNQENVFDLQWME 164  
Qy 177 YFUSEAEDEGLQQRCLDLERQLMLLSEKQS---LAQENAGL-----RERMGRPEGEG 229  
Db 165 VTDMSQEDIE-PLLNKNAHLKRLIDRDEHSETIELSEEDGLHFLPHASSAQSPCG 223  
Qy 230 TPLG-----TAKKULLLOSLEQLEQENFRLESREDERLCALEREVA 274  
Db 224 SPGMKRTESRQHLSVELADAKAKIRRLRQLEEKTEQ--LLDCKQELE-----QMEIELK 276  
Qy 275 ELQHRNALTSLAQEAQKALDEMELROSSRAGOLEATLSCRRRLGELRELROVROL 334  
Db 277 RLOQENMLLSDARSARMYRDELDALREKAVRVKLESEVSRKRLHDIEFYKARVEEL 336  
Qy 335 EERNAGHAERTROLEDEL--RAGS-----LRAQL-----EQRQVQVQELQ 373  
Db 337 KEDNQVLTETMLDQLEGTARSQDKLHELEKENLQKALHDMEMERDMRKKIEELM 396  
Qy 374 QORQEAMKAKWLFECNLEKESVTKERLLAERDSL-REANE-----ELRCAQLQP 428  
Db 337 EENMTLEMAQKQMSDESLHGLWELEQISRTSELSEAPQKSLGHEVNELTSSRLKLEMEN 456  
Qy 429 RGLTQADPSLQPTSTPVDNLAETLPAELRETLLRLQLENKRL----- 471  
Db 457 QSLFTVVELTIVDSVGNASKI-----LKMENQRLSKVLEINEIVQEKQS 507  
Qy 472 ---CROEAAD-----REQEELQRLHLEDANRARGLETOHRLNQOQLSEL 513  
Db 508 LQNCNLKDLKMKKAQLEKTITLRENSERQIKLEQEN--EHLNQVSSLRQSRQISA 565  
Qy 514 RAQVEDIQ---KALQEQGKTEDAISL-----LKRLEELHQLKHEADLELQKREYI 564  
Db 566 EARVQDIEKNILHESIKETSSKLSKTEFEKQRIKKEHYKEGERAE-ELENELHHL 624  
Qy 565 ELEPPTDSS-----TARRIEELQHNQKQADADLAMEERYRYVDKARVMVMOTMEPKQ 618  
Db 625 EKENELQKKTINLKITEKTEA---LEQENSELERENRKLKKTLDLDFKNTLQLESLE 680  
Qy 619 RPAAGAPPELHSLR---TOLRERDVRIRHLEMDPKSRQREQEKK 661  
Db 681 KENSQLEENLELRNRVESLKCASKMAQLQLENKELESEKEQLKK 726

RESULT 7  
AAW79504  
ID AAW79504 standard; Protein; 931 AA.  
XX AC AAW79504:  
XX DT 06-NOV-2001 (first entry)  
XX DE Human protein SEQ ID NO 3150.  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX OS Homo sapiens.  
XX PN WO200157190-A2.  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US04098.  
XX PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52637.  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
useful in diagnosis and gene therapy -  
XX Claim 20; Page 266-267; 6221pp; English.  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAW80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX Sequence 931 AA;  
Qy Query Match 10.8%; Score 391.5; DB 22; Length 931;  
Best Local Similarity 23.6%; Pred. No. 1.6e-19;  
Matches 166; Conservative 143; Mismatches 249; Indels 145; Gaps 26;  
Qy 67 LKVNLSKWLRSVLSVSDVLAHPVSEHLPDVSIGLGE--FSDPA--ELGKLLQVLQCA 122  
Db 2 LRHMNLSLVROIKFYQETLQQLIMMS-LPNVLIIIGKPNFSEQTEEVKKLLLLGCA 60  
Qy 123 ISCEKQDHQRIIMTLEESVQHVMEAIQELMTKDTDPDSLSPETVGN--FDSQSRYYF 179  
Db 61 VOCQKKEEFIERIOGLDFTDRAAANAHIQEV-----THNQENVFDLQWMEYTD 108  
Qy 180 LSEAEAEGLQQRCLDLERQLMLLSEKQS---LAQENAGL-----RERMGRPEGEGTGP 232  
Db 109 MSQEDIE-PLLNKNAHLKRLIDRDEHSETIELSEEDGLHFLPHASSAQSPCGSPG 167  
Qy 233 L-----TAKKULLLOSLEQLEQENFRLESREDERLCALEREVAELQ 277  
Db 168 MKRTESRQHLSVELADAKAKIRRLRQLEEKTEQ--LLDCKQELE-----QMEIELKRLQ 220  
Qy 278 HRNQALTSIAQEAQKALDEMELROSSRAGOLEATLSCRRRLGELRELROVROLER 337  
Db 221 QENMMLLSDARSARMYRDELDALREKAVRVKLESEVSRKRLHDIEFYKARVELKED 280  
Qy 338 NAGHAERTROLEDEL--RAGS-----LRAQL-----EQRQVQVQELQQR 376  
Db 281 NOVLLETETMLDQLEGTARSQDKLHELEKENLQKALHDMEMERDMRKKIEELMEEN 340  
Qy 377 QEEAMKAKWLFECNLEKESVTKERLLAERDSL-REANE-----ELRCAQLQPRGL 431  
Db 341 MTEMAQKQMSDESLHGLWELEQISRTSELSEAPQKSLGHEVNELTSSRLKLEMENQSL 400  
Qy 432 TQADPSLQPTSTPVDNLAETLPAELRETLLRLQLENKRL----- 471

Db 401 TKTVEBLRTVDSVEGNASKI-----LKMKEKNQRLSKKVEILENEIVQEKSLQN 451  
QY 472 CROEAD-----RERQEEQLRHELDANRARGHLETOHRLNQOQLSELRAQ 516  
Db 452 CNLSKDLMKKAQLEKTTETRENSERQIKILEQEN--EHLNQTVSSQURQSISAER 509  
QY 517 VEDLO---KALQEOGKGTDAISIL-----LKRKLEELHQLKHEADLELQKRREYIEL 567  
Db 510 VKDIEKENILHESIKETSSKSKIEFEKROIKKLEHVKGERAE-ELENELHLEKE 568  
QY 568 EPTDSS-----TARRIELOHNLQKDDADLRAMEERYRVVDKARVMQWMEPKORPA 621  
Db 569 NELLQKTKNLKAITCEKIDA-----LEQENSELERENKRLKTYLDFSKNLTFOLESLEKEN 624  
QY 622 ACAPPELHSLR---TQLRERDVIRHLENDPEKRSOREQEK 661  
Db 625 SOLDENLELRNVESLCKASKMAQLQLENKELSEKEQLKK 667  
RESULT 8  
ID ABB12215 standard; peptide; 169 AA.  
XX AC ABB12215;  
XX DT 11-JAN-2002 (first entry)  
XX DE Human hook2 protein homologue, SEQ ID NO:2585.  
XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
KW antifungal; vulenafary; antiulcer.  
XX OS Homo sapiens.  
XX PN WO200157188-A2.  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US03800.  
XX PR 03-FEB-2000; 2000US-0496914.  
XX PR 27-APR-2000; 2000US-0560875.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Liu C, Drmanac RT;  
XX WIPI; 2001-457740/49.  
XX DR N-PSDB; ABA09459.  
XX PT Human proteins and DNA encoding sequences useful for preventing,  
PT treating or ameliorating a medical condition in a mammalian subject  
XX PT e.g. arthritis and cancer.  
XX PS Claim 20; Page 316; 1963pp; English.  
XX CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides,  
CC or polypeptides in a sample, and methods of identifying compounds which

CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.  
CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
CC polypeptide of the invention.  
XX SQ Sequence 169 AA;

Query Match 10.4%; Score 380; DB 22; Length 169;  
Best Local Similarity 97.2%; Pred. No. 1.2e-19;  
Matches 69; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVDKAEKCGSLLTWLTQTHVSPCASPDLSGLAVAYVLNQIDPSWFNEAWLOGISED 60  
|||||  
Db 6 MSVDKAEKCGSLLTWLTQTHVSPCASPDLSGLAVAYVLNQIDPSWFNEAWLOGISED 65  
|||||

QY 61 PGPWNKLVSN 71  
|||||

Db 66 PGPWNKLVTS 76  
|||||

RESULT 9

ABB57920

ID ABB57920 standard; Protein; 1381 AA.

XX AC ABB57920;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 552.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEXE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;  
DR WPI; 2001-658660/75.  
XX N-PSDB; ABL02023.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions  
XX  
XX Disclosure; SEQ ID NO 552; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1381 AA;  
XX  
XX Query Match 10.4%; Score 378; DB 22; Length 1381;  
XX Best Local Similarity 22.3%; Pred. No. 2.6e-18;  
XX Matches 178; Conservative 145; Mismatches 321; Indels 154; Gaps 31;  
XX  
XX 1 MSVDKALCGSLTLLQTFHVPSPCASQDLSGLAVAYLVNLDIPSWFNELAQIGISED 60  
DB 10 WEIDEF-VNSSIIHWLES-----CLPRAELLAGYT-----SLLDGHIHSHWLQ---ID 54  
XX  
XX 61 PGP-----NWK-----LKVNLMKWLRLSVLSYSDVLAHPVSEHLPDVSIGRFS 106  
DB 55 PPQNNPSELNDLNGKLSLSIARAKNFECVVRNKLKSFEEELGQTLV--LPDAFTIGHHP 112  
XX  
XX 107 DP-----AELGKLLQVLGCAISCEKKODHIOIMTLEESVQHVMEATOE-----LMT 155  
DB 113 ESKNGLEQKMTLLTLLGAAVQCPNKELFIARIKELDLTQHAIVGLIKQVTDHSLVLT 172  
XX  
XX 156 KTOPSLSPETGNGFDSQRRYFSEAEDEGLQORCLDLERQMLLSEKQSLAQEN 215  
DB 173 EDSLELTPQS-----WTHILRLTKERDVYKWKID-----LACVETEMTASN 217  
XX  
XX 216 ----AGLRMRGPEGTPGLT-----AKKLLLSQLEQLQENF 253  
DB 218 LVECGGVSVTRSPSGTATSTPSSSSNESNHLAVECADLRKKNRKLQLEKSENLL 277  
XX  
XX 254 RLESREDELRCAELEREVAELQHRNQALTSIAQEAQALKDEMDELROSSERAGOLEAT 313  
DB 278 ELREELDDKKARFDKLQE-----SQWFTEAKRASAYRDEVDILRERAEADRLEVE 330  
XX  
XX 314 LTRCRRRLGELRELROVRLQERNAGHARTQLEDELRRAGSLRAQLEAQRQVQELQ 373  
DB 331 VKRYREKLGDSOFYKSRVRELNRDNLVLESKEMLEQLQ-----Y 372  
XX  
XX 374 GQOEEMAKAEKWLFCERNLEKYESVTEKELLAERDSLRANEELCAQLQPLGLTQ 433  
DB 373 RKRSEHAISLES---BIIKYKKINDMALERD---VDRSKLEELLENSQLQVARNLNS 426  
XX  
XX 434 A---DPSLOPTSTPV---DNLAELIIPALRETLRLLOLENKRL---CROFAADREQEEL 485  
DB 427 TMDLKSFSNEDDNGNSQSEQLTNNQTRALKLELENRLTAAELQKSESPHST 486  
XX  
XX 486 ORHLE-DANRARGLE-TQRLNQOQLSELRAQVEDLOKALOBQOGKTEDAISILKRLK 543  
DB 487 SKMLEKEKKLSLIEQMOENINLTQONVELEGVFNKALEENKLODAVDNRKQSD 546  
XX  
XX 544 EEHLQKHEADLE-LORKREYIEEELPPDSDSTARIEELQHLQKQADLRAEE----- 598  
DB 547 QSLER--EADRKLSDAEQHVETL-----NKEQRIQTILNESIQRRADDLERLAESKTK 599

QY 599 RYRYVDKARVMOT-----MEPKQRPAAAGAPPHELHSLRTQLRERDVR-----IRH 644  
DB 600 ELEYLEKSRQYELTKOKLYEIEARVSTYERENASLLKEVSKLKESQKSVQLDDSDINR 659  
XX  
QY 645 LEMDFE-----KRSQREQ-EKKLLISAWYNGMALOORAGEERAPAHQSFLLAQORLA 697  
DB 660 LDVQSKELQKLGKALEDESEVHQKLEVEKQKQELA-SORIIDQEMISTLRNDLVGTGLV 718  
XX  
QY 698 TNSRRGPLGLRLASLNLRP 715  
DB 719 TKKVRNNLEKLGLADEEP 736  
XX  
XX RESULT 10  
XX ABG07202  
XX ID ABG07202 standard; Protein: 320 AA.  
XX AC ABG07202;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #7193.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO2001/75067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Dmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS71389.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity  
XX  
XX Claim 20; SEQ ID No 37561; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes;  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.









CC anaemia, platelet disorders, thrombocytopaenia, wounds, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.

XX Sequence 1879. AA;

Query Match 8.9%; Score 322; DB 22; Length 1879;

Best Local Similarity 21.1%; Pred. No. 4.1e-14;

Matches 136; Conservative 132; Mismatches 254; Indels 204; Gaps 29;

QY 126 EKQDHIQIMTLEESVQH-----VYMEAIQELMTKDTFDSLSPTFYGNFDSOSRYIFL 180

DB 1095 EMKQHAQVQSLEQEQKRAKANLKNKOTLEKENTD-LAGELRVLGQAQKEVEHRM 1153

QY 181 SEAESEDELOQCILDERQLMLLSEKQSLAQENAGLRERMGPRGEGTGLTAKKLLL 240

DB 1154 KFLQAQVQELQSCDGERARAEINLQNEVESVTGMLNEAGKAIK--LAKDVAS 1211

QY 241 LQSOLEQOENFRLESGBREDELRCAELEREVAELQHRNOALTSIAQEAQALKDEM-- 298

DB 1212 LSSQLQDQE-----LLOESRQKLNVTSLRLEEEERNSLDQDLEE 1254

QY 299 -ELRQSSER-AGOLEATLTCRRRLGEL-----RELRRQVRQLEERNAGH 341

DB 1255 MEAKQNLERHISTNLQDSKKKLODFASTVEALEEGKRFQKEIENITQOYEERAAAY 1314

QY 342 --NERTP-QLEDELRLRAGSLRAQLEORRQOVQLOGORQOEAMKAELWPECNLEEKY- 397

DB 1315 DKLEKTKNRLOQEL---DDLVDLDNQRLVSNLEKKQ---KFDOLLAEBEKNISKYA 1367

QY 398 -----ESVTKERE-----RLLAEE-----RDSLRANEELRCQAQLQPRGLTQ----- 433

DB 1368 DERDRVEAREKETATSLARALEALEAKEBELERTNKMKAEMGPGSKDDVQOEL 1427

QY 434 -----ADPSLDPTSTPDVNDLAAEILPAELRETLRLRLQLENKRLCRQE-----AA 477

DB 1428 SHDLEKSKRALGDPRLPEMTQLEELGRTELASPRDQAKLRLEVNMQAPSRASFERDLQA 1487

QY 478 DRROEELQRLHEDANRARGLETO---HRLNQ-----HRLNQ----- 507

DB 1488 RTEQNEESRRHLQ---ROLHEYTELEDERQKRALAAAKIKLGWDVPVRLDLXADSAIK 1544

QY 508 -----QQLSELRAQVEDLQALQEQGGTDAISILK-----RKLEEHLOKLHE--- 552

DB 1545 GRGKAIKQLKQAKQKDFQRELEDRASDEIFATAKENEKAKSLEADLMQLEDLA 1604

QY 553 -----ADLEQKREYIELEPPPTDSSTA-----RRIEELQHLNLOKDDADLRA 595

DB 1605 AAEGRQADLE---KEELAEELASSLGRNALQDEKRRLEARTALQEELEEEBQGNMEA 1661

QY 596 MEERYRYVDKARWMTMEPKORPA---AGAPPELH-----SLTQLRER----- 638

DB 1662 MSDRVKATQAEQLSRELATERSTAQNESARQOOLERQNKELRSKLHEMAGVYKSKFKS 1721

QY 639 -----DVRIRHLEMDFEK-----SRSQREQEKL---LISAWYNGMALQORACEER 682

DB 1722 TIALEAKIAQLEQVQEAEREKQAATSLKQDKKLKEILLQVEDEKRMKAEIQEAK 1781

QY 683 APAHAQSFQAO-----QRLATNSRR 702

DB 1782 GNARVQKLRQLEAEESQINANRRK 1809

RESULT 15

AAB69070

ID AAB69070 standard; Protein: 1374 AA.

XX AAB69070;

AC AAB69070;

DT 19-APR-2001 (first entry)

XX

Human male enhanced antigen-2 (MEA-2) protein sequence SEQ ID NO:2.  
Human; male enhanced antigen-2; MEA-2; identification; spermatogenesis;  
spermatogenesis disease; chromosome marker; pancreatic cancer.

OS Homo sapiens.

XX JP2000316590-A.

XX 21-NOV-2000.

XX 30-APR-1999; 99JP-0125196.

XX 30-APR-1999; 99JP-0125196.

XX (ITOH-) ITO HAM KK.

XX WPI: 2001-128256/14.

XX N-PSDB; AAF32508.

XX A new protein, human male-enhanced antigen-2, useful for detecting

spermatogenesis diseases

XX Claim 1; Page 12-15; 21pp; Japanese.

XX The present sequence represents the human male enhanced antigen-2

(MEA-2). The present invention also described an antibody specific for  
the MEA-2 protein. The antibody can be used for the identification of a  
gene causing diseases related to spermatogenesis. The MEA-2 nucleotide  
sequence is useful as a chromosome marker, and in the detection of  
pancreatic cancer.

XX Sequence 1374 AA;

Query Match 8.8%; Score 319; DB 22; Length 1374;

Best Local Similarity 24.0%; Pred. No. 4.5e-14;

Matches 187; Conservative 129; Mismatches 286; Indels 176; Gaps 33;

QY 66 KLVSNLKM-VLRSVYSQDVLAHPVSEHLFDVSLIGFSPAEGLKLLQLVLGCAIS 124

DB 544 KLASQAEISSLSQSVRWYQQLA--LAQE--ARVRLQGEWAH-IQVGQMTQAGLLEHLK 598

QY 125 CEKKQDHIQIMTLEESV-----QHVMEAIQELMTKDTFDSLSPE 165

DB 599 LENVLSQQLTEQHRSMKEKGRIAAQLOGIEADMLDQEAFAFMQIQEAKTMEED--LQR 656

QY 166 TYGNFDSQSRYYFLSEEA-----EEGDELQORCLDLERQLMLLSEE 207

DB 657 RLEEFGERERLQRMADSAASLEQOQVKTLLQROQOQLEAQOEHLDLMKQLTLTQEA 716

QY 208 KQSLAQENAG-----LRERMGRPEGE-----TPGLTAKKLLLSQSL-----EQL 248

DB 717 LOSREQSIDALQTHYDELOARIGELQEAASREDTICLLONEKIILEAALQAQKSGREL 776

QY 249 QENFRLESRED-----ERLRCALEREVAELQHRNOALTSIAQEAQALKDE----- 296

DB 777 DRGARREEGTEETSETLEKL--BELAIKSGQVEHLAQOETAALKKQMKIKEQFLOOKVM 835

QY 297 MDELQSSERAGOLEATLTSORREL--GELRELROVRQLEERNAGHAERTROLEDELRA 355

DB 836 VEAYRRDATSKDQLISELKATRKRLDSELKELRELQMQV-----HGK-RTAAELSR- 887

QY 356 GSLRAQLEAORROVOELQGOQOEAMKAELKLFECRLNEEKYESVTKEKELLAERDSL 415

DB 888 --LHREVAQVRQHMDLEGLHQ-----SAQK---ERDEMETHLSQLOQPDKEQWA---VT 934

QY 416 EANEELR--CAQLQPRG-----LTQADPSLDPTSTFVDNLAELIPAE 456

DB 935 EANEALKQIEELQEAQKAITQOKMRRIGSDITSQAQEMKTKKRAYEN-AVGILSR 993

QY 457 LRETLRLQLENKRLCRQEA-----ADREQBELQRLHEDANRARGHLEGTQHR--- 504

DB 994 LQEAALAAKAAADAEIGQLRAQGGSSSSSLALHERIQALEAQVSHSKTLLLEKLEQVI 1053

Search completed: March 4, 2003, 14:58:45  
Job time : 71.5679 secs.

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:52 ; Search time 32.3446 seconds

(without alignments)  
2137.006 Million cell updates/sec

Title: US-09-763-902B-5

Perfect score: 3638

Sequence: 1 MSVDRAELCGSLTLWLOTFH.....SRRGPLGLASLNLRPTDKH 719

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

PIR\_73:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 360   | 9.9         | 1407   | 1 S28589 | trichohyalin - rab |
| 2          | 344   | 9.5         | 4684   | 2 A59404 | plectin [imported] |
| 3          | 343.5 | 9.4         | 4574   | 2 G02520 | plectin - human    |
| 4          | 342.5 | 9.4         | 2442   | 2 T08621 | centrosome associa |
| 5          | 342.5 | 9.1         | 4687   | 1 A39638 | plectin - rat      |
| 6          | 329.5 | 9.1         | 1909   | 2 A45592 | liver stage antige |
| 7          | 326.5 | 9.0         | 772    | 2 T27307 | hypothetical prote |
| 8          | 322   | 8.9         | 1999   | 1 S21801 | myosin heavy chain |
| 9          | 319   | 8.8         | 1549   | 1 A40691 | trichohyalin - she |
| 10         | 316   | 8.7         | 1961   | 1 A61231 | myosin heavy chain |
| 11         | 315.5 | 8.7         | 1201   | 2 B35815 | myosin heavy chain |
| 12         | 315   | 8.7         | 1898   | 1 A45973 | trichohyalin - hum |
| 13         | 314.5 | 8.6         | 1175   | 2 D35815 | myosin heavy chain |
| 14         | 314.5 | 8.6         | 1956   | 2 T16416 | hypothetical prote |
| 15         | 312   | 8.6         | 1938   | 1 A40997 | myosin heavy chain |
| 16         | 311.5 | 8.6         | 2101   | 2 A42184 | nuclear mitotic ap |
| 17         | 309.5 | 8.5         | 2411   | 2 B32491 | myosin heavy chain |
| 18         | 309   | 8.5         | 1992   | 2 A47297 | myosin heavy chain |
| 19         | 308.5 | 8.5         | 1218   | 2 T14265 | golgin-245 - mouse |
| 20         | 308.5 | 8.5         | 2385   | 2 A32491 | myosin heavy chain |
| 21         | 307   | 8.4         | 746    | 2 T47237 | myosin II heavy ch |
| 22         | 306.5 | 8.4         | 1976   | 2 A59252 | myosin heavy chain |
| 23         | 306   | 8.4         | 886    | 2 H69378 | conserved hypothet |
| 24         | 304.5 | 8.4         | 1426   | 2 T00337 | hypothetical prote |
| 25         | 304   | 8.4         | 2007   | 1 B43402 | myosin heavy chain |
| 26         | 302.5 | 8.3         | 1039   | 2 S18199 | myosin heavy chain |
| 27         | 302.5 | 8.3         | 1931   | 2 A59234 | slow myosin heavy  |
| 28         | 302   | 8.3         | 978    | 2 A70387 | conserved hypothet |
| 29         | 301   | 8.3         | 955    | 2 S24348 | myosin heavy chain |

30 300 8.2 1974 2 T30010  
31 299.5 8.2 1827 2 T16270  
32 299 8.2 1133 2 T22976  
33 298.5 8.2 1940 2 A59287  
34 298 8.2 1964 2 A59282  
35 297.5 8.2 1201 2 A35815  
36 297.5 8.2 1959 1 A33977  
37 297 8.2 857 2 S33821  
38 296.5 8.2 1175 2 C35815  
39 296 8.1 1972 1 A41804  
40 295 8.1 1388 2 S70533  
41 295 8.1 1388 2 S74245  
42 294 8.1 924 2 S06117  
43 294 8.1 936 2 S39083  
44 293 8.1 1738 2 T14867  
45 292 8.0 1790 2 S67593

## ALIGNMENTS

## RESULT 1

S28589

trichohyalin - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 12-Mar-1993 #sequence\_revision 01-Mar-1996 #text\_change 22-Jun-1999

C:Accession: S28589

R:Flitz, M.J.; Rogers, G.E.

submitted to the EMBL Data Library, December 1992

A:Description: Examination of the gene encoding rabbit trichohyalin.

A:Reference number: S28589

A:Accession: S28589

A:Molecule type: DNA

A:Residues: 1-1407 &lt;FIE&gt;

A:Cross-references: EMBL:Z19092; NID:g1746; PIDN:CAA79519.1; PID:g1747

C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root

Covalent modifications to this: protein include conversion of arginine to citrulline a

C:Genetics: 46/3

A:Introns: 46/3

C:Superfamily: trichohyalin; calmodulin repeat homology

C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F:49-81/Domain: calmodulin repeat homology &lt;EF2&gt;

Query Match 9.9% Score 360 DB 1 Length 1407

Best Local Similarity 25.5% Pred. No. 1.9e-08

Matches 164; Conservative 110; Mismatches 262; Indels 108; Gaps 21

Qy 129 QDHIQIRIMTLEESVQHVYVNEAIQELMTKDTDPDSLSPTETGYNFDSQSRYYFLSEAE--E 186

Db 168 RDKEQRLQORQLEERRAEEREEQRLRRKGRDAEEFEEELRRRQEQELKRELREERQORRE 227

Qy 187 GDELOQRCIDLERQIMLLS-----EKKSLAQNAGLRNMGPEGE--GTPGLTAKK 237

Db 228 RREHERALQEEEOQLLRQRWRREPREQQLRRELEIREERQLEERQLEERQLEERQLEERQ 287

Qy 238 LL-----LLOSOLEQLOENFELESGR-----EDERLCARLEAEVLAELQHNQAL 283

Db 288 RLEQERREQQLRRELEIREERQLEERQLEERQLEERQLEERQLEERQLEERQLEERQ 347

Qy 284 TSAQEAQALKDENDE-----LRQSSERAGQLEATLTSCRRRLGELRELRRQ 330

Db 348 EQERERREQLLAEEVREQARERGESLTRWQQLSEAGARQSKVYSRPRQEE-----QS 402

Qy 331 VRQLEERNAGHAETROLEDELRLRAGSLRA-----OLEAQRORQV 370

Db 403 LRQDQERQ-RQERERELEEQARQQQQAEESEERRRQLRSARPSRLRQRAERQEQ 461

Qy 371 ELQQRQEAEMKAEKWLFCRNLEKYESVYKEKRLLAERDSLREANDELCAQLQPRG 430

Db 462 E-QRPREERQORRR-RQELQFLEEEQLQRRAQQQLEEDSFQEDRRRRRQEQPRG 519

Qy 431 LTQADPSLDPTSTPVDNLAAILPAELRETLRLQLENKRLCQEAADRQEE--LORH 488

Db 520 QTRWQLOEQARRRHTLYAK--PQOQEQ--LREBEELQREKKRQEREREYREBEKLORE 575  
Qy 489 LEDANRARGHLETOHRLNQOQLSELRAVEDLQKALQOQGGKTEDAISILK-----RKL 543  
Db 576 -EDEKRRQEREROYR-----ELEELQOEBQDRKRLREEQLOOERERLRQERERKL 630  
Qy 544 EEHLQKHEADLELQKREYIELEPPDSTARRIEBLQHLNQLKQKADLRAEMEYRY 603  
Db 631 REEBQLOQOEBQLOEQRE-----RKLREEQLOOERERLRQERERKLREEQLO 674  
Qy 604 VDKAMVMQTEPKORPAAGAPPELHSLRTOLRERDVIRHLEMDFEKRSQREQEKLL 663  
Db 675 REEBQLOQOEBQLOEQRE-----RKLREEQLOOERERLRQERERKLREEQLO 726  
Qy 664 ISAWYNGMALQOQAGEERAPAHQAQSLAQOQLATNSRGPGLR 707  
Db 727 LRR--EOLLQOERDKLREBEQLOOQSEERLRQREQOQLRR 768

RESULT 2  
A59404  
plectin [Imported] - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 19-Apr-2002  
C:Accession: C59404; A59404  
R:Li, C.G.; Maercker, C.; Castanon, M.J.; Hauptmann, R.; Wiche, G.  
Proc. Natl. Acad. Sci. U.S.A. 93, 4278-4283, 1996  
A:Title: Human plectin: organization of the gene, sequence analysis, and chromosome localization  
A:Reference number: C59404; MUID:96210632; PMID:8633055  
A:Accession: C59404  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-4684 <STO>  
A:Cross-references: GB:CAA91196; NID:gl296662; PIDN:CAA91196.1  
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S

Query Match 9.5%; Score 344; DB 2; Length 4684;  
Best Local Similarity 24.2%; Pred. No. 3.5e-07;  
Matches 196; Conservative 109; Mismatches 250; Indels 254; Gaps 36;

Qy 65 WKLVSNLKMVLSVYSQDVLAH-PVSEHPLDVS-----LIGFSDPAELGKLQLVL 119  
Db 1314 WREVA-----QLLRWQAVLAQTVROREQLQRLRYRESADP--LGAWLQ--- 1361

Qy 120 GCAISCEKKODHQTORTIMLEESVQHVYMEAIQELMTKDPDLSPTGYNFDQSRRYYF 179  
Db 1362 ----DARRQEQIQ-AMPLADS-----QAVRE-----QLRQEQA 1390

Qy 180 LSEEAEGDELOQRC-----LDLERQLMLLSEE-----KQSLA 212  
Db 1391 LLEEIERHGEKVEECORFAKOYINAKDYELQVLYKAQLEVPASPAKKPKVQSGSESVI 1450

Qy 213 QENAGLRMRGPGEPTGLTAKKLLLOSQLEQNEFRLESGREDELRCAEL--- 269  
Db 1451 QEYVDLTHYSE-----LTTLTQYIKFISETLRMEERERLAQQQAEERERLAEVAA 1505

Qy 270 -----EREVALQ-----HRNQALTSLAQQAQKALDEMDLRQS 303  
Db 1506 LEKOROLAEHAQAQAQAEAKELQRIQOEIVVREAAVDAQQKRSIQEELQRLQS 1565

Qy 304 SE-----RAGQLEATLTSRRRLGELRELRRQVRLQERNAQ----- 340  
Db 1566 SEAEIQAKARQAQAASRLRIEIEIRVRLQLEATERQRGGAEGELQALRAAEAEQAQ 1625

Qy 341 -----HAER--TROLELAR-----AGSLRAQLEAQRROVQELQ--QRQEEAMKA 393  
Db 1626 KROQAEERLRQVQDSQKQQAQAEVLAASRVKAEAAAREKORALQALEELRLQAEAA 1685

Qy 384 EKWLFE-----CRNLEKYESVTKEKRLAERDSLRANEELRCALOPLRGLTOADPSL 438  
Db 1686 ERWLCOAEVERAROVQVLAETQASAE--AELOSKRASFAE-KTAQLE-RSLQEEHVAV 1740

Qy 439 DPTSTPVDNLAAEILPAE-----LRETL-LRLQ-----LENKRLCRQEA- 476

Db 1741 AQLREERARRAQQAQAEARAREAEERQLERWQKANEALRLRLQAEVLOQKSLAQAEAE 1800  
Qy 477 -----ADRE-----ROEEL-ORHLEDANRARGHLETOHRLNQOQLSELRA--- 515  
Db 1801 KQKEAEAREARRRQKAEQAVRORELAPQOELKQRLQAEQTAQOQLAAEQELIRLRAETE 1860  
Qy 516 QVEDLQKALQOQGGKTEDAISILKRLKEELHLOKLHLEADLELQKREYIELEPPDST 575  
Db 1861 QGEQORQLLEELAR-----LQREAAATQROELEAEALAKVRAEMEVLL-----ASK 1908

Qy 576 ARRIEELQHNLOKQKADLRAEMEYRYRYDKA-----RMVMTMEPKQRPAA 622  
Db 1309 AKAEESRSTSEKSKQRLAEAGRFRELAEEAARLALAEAKRQROLAEADAARQAEAA 1968

Qy 623 -----GAPPELHSLRTOLRERDVIRHLEMDFEKRSQREQEKLLISAWYNGMAL 674  
Db 1969 ERVLAELKLAIGEATRLKT---EAEIALKEKEAENELRLRAEDE-----AF 2012

Qy 675 QORAGEERAPAH---AQSFQAQOQLATNS 700  
Db 2013 QRRLEEQAAQHKADIERLAQLRKAQSDS 2041

RESULT 3  
G02520  
plectin - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 20-Aug-1999  
R:McClean, W.H.I.; Smith, F.J.D.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: H01385  
A:Accession: G02520  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4574 <MCCL>  
A:Cross-references: EMBL:U53204; NID:gl1477645; PIDN:AA05427.1; PID:gl1477646  
C:Genetics:  
A:Gene: PLEC1  
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein  
F:68-283/Domain: alpha-actinin actin-binding domain homology <ACT>

Query Match 9.4%; Score 343.5; DB 2; Length 4574;  
Best Local Similarity 23.2%; Pred. No. 3.6e-07;  
Matches 184; Conservative 126; Mismatches 263; Indels 221; Gaps 32;

Qy 67 LKVSNLKMVLSVYSQDVLAHVPSEHPLDVSIG-----EFSDDPAELGKLQLVL 119  
Db 1199 LEVERWRERVAQVLERWQAVLAQ--TDLRQRELEQLQRLRYRESADP--LGAWLQ--- 1251

Qy 120 GCAISCEKKODHQTORTIMLEESVQHVYMEAIQELMTKDPDLSPTGYNFDQSRRYYF 179  
Db 1252 ----DARRQEQIQ-AMPLADS-----QAVRE-----QLRQEQA 1280

Qy 180 LSEEAEGDELOQRC-----LDLERQLMLLSEE-----KQSLA 212  
Db 1281 LLEEIERHGEKVEECORFAKOYINAKDYELQVLYKAQLEVPASPAKKPKVQSGSESVI 1340

Qy 213 QENAGLRMRGPGEPTGLTAKKLLLOSQLEQNEFRLESGREDELRCAEL--- 269  
Db 1341 QEYVDLTHYSE-----LTTLTQYIKFISETLRMEERERLAQQQAEERERLAEVAA 1395

Qy 270 -----EREVALQHRNQ-----ALTSLAQQAQKALDEMDLRQS 303  
Db 1396 LEKOROLAEHAQAQAQAEAKELQRIQOEIVVREAAVDAQQKRSIQEELQRLQS 1455

Qy 304 SE-----RAGQLEATLTSRRRLGELRELRRQVRLQERNAQHAERTROLEDELRLRAGSL 358  
Db 1456 SEAEIQAKARQAQAASRLRIEIEIRVRLQLEATERQRGG-----AEGELQALRAR 1508

Qy 359 RAQLEAQRROVQV-----LOGORQEEAMKAQKWLFECLNLEKYESVTKEKRLAERDSL 414

Db 1509 AEEAQRQAQAEERLRQVQDESQRKQAEVELASRVKAEATAAREKQALQALEEL 1568  
Qy 415 R-----EANEELRCAQLPRGLTQ-----ADPSLDPTTPVDNLAAPAEILPAELRE-- 459  
Db 1569 RLQAEAEERLRQAEVERARQVVALETAQASAEALQSKRASPFAKTAQ--LERSLQEEH 1627  
Qy 460 -TLRLQLENKRLCROEAADREOREEELQHLHLE---DANRA--RHGLETOHRLNQOOLSEL 513  
Db 1628 VAAQLREAEERLRQAEAEERAEERELERWQLKANEALRLRLQAEVVAQKSLAQ 1687  
Qy 514 RA--QVEDLQKALQEOGGKTDAISILLKRLK--BEHLQKLHE-----ADLEL-- 557  
Db 1688 EAEKQEAEREARRRGAEEQAVR---QRELAQEELKQRLAEGTAQRLAAEQELIR 1744  
Qy 558 -----QRKREYIE-----LEPPTSSTARRIEELOHNLQKKDADLRA-- 598  
Db 1745 LRAETEOEQOQRLLEELARLQREAAATQKR--QELEAEELAKVRAEMEVLLASKARAE 1803  
Qy 599 RYRVVDKARVMQTMPEKQAPAGAPPELHSLTQLRERDVRIHLEMDFEKRSQREO 658  
Db 1804 ESRSTSEKSKORLEAAGRFRELAEEAARLALAEAK-----RQRLAEEDAAARQRAE 1857  
Qy 659 EEKLISAWYNG-----MALQQRAGEERAPAH--- 686  
Db 1858 AERVLAELKLAIGEATRLKTEAETALKEAEENELRLAEDEAFQRRLEEQRAHQKAD 1917  
Qy 687 AQSLAQOORLATNS 700  
Db 1918 TEERLAQLRKASDS 1931  
  
RESULT 4  
T08621  
C:Species: human  
C:Species: Homo sapiens (man)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: T08621  
R:MacK, G.J.; Rees, J.; Sandblom, O.; Balczon, R.; Fritzler, M.J.; Rattner, J.B.  
A:Title: Autoantibodies to a group of centrosomal proteins in human autoimmune sera read  
A:Reference number: 216462; MUID:98165428; PMID:9506584  
A:Accession: T08621  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2442 <MAC>  
A:Cross-references: EMBL:AF022655; NID:g2832236; PIDN:AAC06349.1; PID:g2832237  
A:Experimental source: cell line HeLa  
  
Query Match 9.4%; Score 342.5; DB 2; Length 2442;  
Best Local Similarity 23.7%; Pred. No. 2e-07;  
Matches 166; Conservative 114; Mismatches 224; Indels 197; Gaps 28;  
  
Qy 70 SNLKMVRLSVESQDVLAHPVSEEHLPDVSIGFSDPAELGKLLQLVLGCAISCEKKQ 129  
Db 1616 SHSTVLARELQRDQEVKSQRQIEEL-----QRQK 1646  
  
Qy 130 DHI-----QRIMTLESVQHVYMEAIQELMTKTPD-----SLSPETYNFDSQSR 176  
Db 1647 EHLTQDLERRDQELMKQERIQ--VLEDQRTQTKILEDLQIKLSLRGRGRLTQ 1704  
  
Qy 177 YFLESEAEAG-----DELQORCLDLERQLML-----LSEKQSLAENAGL 218  
Db 1705 --LMQRAEEGRGPSKAQSGSEHMKLIILRDKEVEYEQCOEHHELOELKDLQEQLOQL 1762  
  
Qy 219 RERMRPEEGTGGTAKLLLLQSQLEQ-----LQENRFLSGRDEBLRCAELEREVA 274  
Db 1763 HRKVGETS-----LLLSQREQIVVLOQ-----LQEARQGLKEQSLOSOLD 1806  
  
Qy 275 ELQ-----HRNQALTSIAQPAQALKDMDLROSSER--AGOLEATLTSRRRLGELREL 329  
Db 1807 EAQRALAQRDQLEALQEQQAQGGQEEVKKADALQALEGAHMTLKERHGLQDHKE 1866  
  
Qy 330 QVRQLERNAGHARTRQLEDELRLRAGSLRAQLEAQRQVRQVQELQORQOEAMKAEWLFE 389

Db 1667 QARRLEELAVEGRRVQALEVL---GDRAESREQEKALLALQOQCAEQAEHE---VE 1920  
Qy 390 CRNLEEKYESTYKKEKRLAERDSLRANEELRCAQLPRGLTQADSLDPTSTPVDNLA 449  
Db 1921 TRALQDSW-----LQAQVWLKERDQELEA----- 1944  
Qy 450 AEILPAELRETLLRLQLENKRLCROEAADREOREEELQHLHLEMDFEKRSQREO 509  
Db 1945 -----LRAESOSSR--HQEEAARAEALQALQKANALOGKE--QHLLBQAE 1989  
Qy 510 LSE-----LRAQVEDLQKALQEOGGKTDAISILLKRLKLEHKLQHLHEAD 554  
Db 1990 LRSISLEASTATLQASLDACQASHRSQLEALRIQGEIQDQ-----DLRYQEDVQOQOAL 2044  
Qy 555 LELQKREYIELEPPDTSSTARRIEE--LOH-----NLQKKDADLRAEMEYRR--YV 604  
Db 2045 AQRDLELHOOERQLEKSLAQRVQENMIQEKQNLGLEREEREEIHLQHSVRELQTLA 2104  
Qy 605 DKARVMQTMPEKQAPAGAPPELHSLTQLRERDVRIHLEMDFEKRSQREO 664  
Db 2105 QKEOILELRETQORNNLEALP--HSHKTSPEWEE-----QSLKLSLEPLRLQRELE- 2153  
Qy 665 SAWYNGMALQORAG-----EERAPAHQASFLAQORLATNS 700  
Db 2154 ---RLQAALQRTAEARETEWREKAQDLALS--LAQTAKSVSS 2189  
  
RESULT 5  
A39638  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: A39638; S21876  
R:Wiche, G.; Becker, B.; Lubner, K.; Weltzer, G.; Castanon, M.J.; Hauptmann, R.; Strat  
J. Cell Biol. 114, 83-99, 1991  
A:Title: Cloning and sequencing of rat plectin indicates a 466-kD polypeptide chain w  
A:Reference number: A39638; MUID:91268156; PMID:2050743  
A:Accession: A39638  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-4687 <MTC>  
A:Cross-references: EMBL:X59601; NID:g1292885; PIDN:CAA42169.1; PID:g1561642  
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protei  
C:Keywords: cytoskeleton; transmembrane protein  
F:6-103/Domain: ribosomal protein S10 homology <RS10>  
F:184-399/Domain: alpha-actinin actin-binding domain homology <ACT>  
  
Query Match 9.1%; Score 332.5; DB 1; Length 4687;  
Best Local Similarity 22.2%; Pred. No. 1.1e-06;  
Matches 193; Conservative 117; Mismatches 272; Indels 287; Gaps 32;  
  
Qy 68 KVSNLKMVRLSVESQDVLAHPVSEEHLPDV----- 99  
Db 1227 KLKTSILVIRS--TQGAEEVL--KTHEEHLKEAQAVPATLQLEVTKASLKLRAQAEAAQ 1283  
  
Qy 100 -----SLIGFSDPAELGKLLQLVLGCAISCEKKQDHQIRIMTLESVQHV--VMEAIQEL 153  
Db 1284 PVFNTLRDLRGAQAEYGERLQ-----QRGERDVEVERWRERVOTLLERWQAV 1331  
  
Qy 154 MT-----KDTPLDSLPETYNFDSQSRYYF----- 179  
Db 1332 LAQTDVQRQELRQGLRQRYRESADPLSSWLO---DAKSQEQIQAVPIANSQAAREOL 1388  
  
Qy 180 -----LSEBAEGDELEQRC-----LDLERQLMLSEE----- 207  
Db 1389 ROEKALLEEIERHGEKVEECQKFAQYINAIKDYELQITYKAYQLEPVASPAKPKVQSG 1448  
  
Qy 208 KOSLAQENAGLRMRGPEGTPGLTAKKLLLLQSLQLEQLEENFRLESGRDELRCA 267  
Db 1449 SESVIOEYVDLTRYSE-----LTTILTQYIKFISETLRMEEEERLAEQQAERERLA 1503  
  
Qy 268 ELER-----EVAELQHRNQ-----ALTSLAQEAQALKDEND 298



A:Reference number: S34842  
A:Accession: S34842  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 323-381, 'HKAI' <GUE2>  
A:Cross-references: EMBL:N28266  
A:Note: difference at carboxyl end due to frameshift error  
C:Comment: This protein is found as flocculent material in the parasitophorous vacuol  
C:Superfamily: trichoyallin; calmodulin repeat homology  
C:Keywords: EF hand  
F:154-1629/Region: 17-residue repeats (A-K-E-R-L-Q-E-Q-Q-S-D-L-E-Q-E-R-R)

| Query Match           | 9.1%              | Score 329.5;       | DB 2;       | Length 1909; |
|-----------------------|-------------------|--------------------|-------------|--------------|
| Best Local Similarity | 24.5%             | Pred. NO. 5.6e-07; |             |              |
| Matches 155;          | Conservative 125; | Mismatches 236;    | Indels 120; | Gaps 27;     |

[illegible]

QY . 289 EAQALKDEMDELQSSERAGQLEATLTSCRR-----RLGELRELRRQVRQLEERNAGH 341

432 EQQS--DLEQLAKEKLQEQQSDLEQERRAKEKLQEQQS-D 487

342 AERTRQLEDELRRAQSLRAQLEAQRQVQVQELQGRQEEAMKAQKWLFEERNLEEKYESVT 401

db 488 LEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDSEQLAKEKLQEQQSDLEQ--ERLA 545

402 KEK-----ERLLAERDSLREANEELRCAQLQPRGLTQADPSLDPTSTPVDNLAAE 451

546 KEKLEQQSDLEQERLAKEK--LQEQSDLEQERLAKEKLEQEQSDLEQ-----ERLAKE 598

QY 452 · ILPAELRETLRLQLENKRLCR-----QEAADRERQEELQRHLEDANRARHGLE 500

db 599 KLQKQ-----QSDLEQERLAKEKLQEQSDLEQERLAKEKLQ 652

QY 501 TQHRNQQLSELRAQVEDLQ-----KALQEQGKTEDAISILLKRKLEEHL 547

db 653 EQ---QSDLERTKASKETLQEQQSDLEQERLAKEKLQEQQSDLEQ-----ERRAKEKL 702

QY 548 QKLHEADLELQK-REYIEELEPPTDSSTARRIEELQHNLOKKDADLRAMEERYRRYVDK 606

Db. 703 QE-QQSDLEQERRAKEKLEQQ--SDLEQERRAKE---KLQEQQSDL---EQERRAKEK 752

607 ARMVMTMEPKORPAAGAPPELHSLRTLRLERDVRIRHLEMDFEKRSQRE--QEEKLLI 664

db 753 LQEQQSDLE-QDRLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQDRLAKEKLQEQQSDL 811

QY 665 SAWYNMGMALQQRAGEERAPAHQAQSF LAQQR LA 697

812 E-----QERRAKEKL-QEQQSDLEQERLA 834

T27907

C; Species: *Caenorhabditis elegans*

Sequence\_revision 13 OCT 1999 #text\_change 13 OCT 1999  
C;Accession: T27907

submitted to the EMBL Data Library, June 1995

A; Reference number: Z20437

A;Status: preliminary; translated from GB/EMBL/DDBJ

A: Molecule type: DNA  
A: Residues: 1-772 <HAL>  
A: Cross-references: EMBL:U29380; PIDN:AAA68733.1; CESP:2K546.1  
A: Experimental source: strain Bristol N2  
C: Genetics:  
A: Gene: CESP:2K546.1  
A: Introns: 53/3; 83/2; 477/1; 496/2; 632/3; 692/3; 729/2; 745/2

Query Match 9.0%; Score 326.5; DB 2; Length 772;  
Best Local Similarity 21.1%; Pred. No. 2.9e-07;  
Matches 164; Conservative 153; Mismatches 320; Indels 141; Gaps 27;

QY 1 MSVDKAEICGSLTTLWQTHVPS-PCASPDLSGLAVAYVNLQIDPSWFNANLQIGSE 59  
DB 43 LOADLADMA---VMEGLDATKLPNDPQLCNGRAFSEVLHNDKFNFTDGLWETMPE 98  
QY 60 DPGPNKLVKSNLKVLSVYSQDLAHPS-----EHLDPVSLIG-E 104  
DB 99 NRTTNVFRSTRKLRKMFYVNHINTVSSRWTDIHERIDGIYESDLPAVNLGNA 158  
QY 105 FSDPAELAK----LLOLVGCAISCEKKODHRIWT---LEESVQHVVMETAEOLMTKDT 158  
DB 159 VVTLAHIGNAKRFVDYSKALTSTHKSMNSAKVMTVVIDEMPENPCFHEISELHGSO 218  
QY 159 PDSLSPTYGNFDSQSRYYFLSEAEDEGLQRCILDLERQLMLLSEKQSLAQENAGL 218  
DB 219 ELNLSSESGKLNGS-----SERRSNADQI---LVDAEIEIRLRTETENORKEIERL 270  
QY 219 RRMGRPEGPGPGLTAKLLQLQLEQOENFRLESGREDRLRCACELREVAELQH 278  
DB 271 TKSFTAQHDMSSNESGDISILEKQNEELRKRLEKNEKLEDAVDQFGVIFELTN 330  
QY 279 RNOALTSIAQEAQALKDEM-----ELROSSERAGOLEATLTSCRRRLGRELRLROVROL 334  
DB 331 ENDVLRSDKERQRLQTVLDAQSDLDKWTYQVQKEALSQKQDKKEIKELLSQNKAL 390  
QY 335 ERNAGHAERTQLED-----ELR-RAG-----SLRAQLEAQRQVQELQORQEE 379  
DB 391 KSR-LDHVYKATLEDANKNGIAQLRTQVGGTALNTLKSLSKRCVQELTQLOH 449  
QY 380 AMKAEKWLFECLNEKESVYKKEKRLAERLAEERLREANELRCALQPRGLTQADPSLO 439  
DB 450 -----REKVKELDRKDELTEENRLE--NQLIFKEAVTP-----SLH 486  
QY 440 PTSTPDVNLAEILPAELRETLRLQLENKRLCROSPADRERQEEQLRHLEDANRAHGL 499  
DB 487 ESMFEAGNLS-----EKNTLPLEIENKRL-TERIQELESLEPLKGLITLTKSNGVL 537  
QY 500 ETQHLNQOQLSELRAQVEDLQK-ALQEOGGKTEDAISILLK-RKLEEHQKLHLEADLEL 557  
DB 538 EBEKLFATKQIEELQOQIEDLQENLLKNOEHASGDVYGLKIQLEKAEVAEQQMREAKRA 597  
QY 558 QKREYIELEPPTDSSTARRIEELQHNKKDADLRAEMEYRRYVVDKARMVMOTMEPK 617  
DB 598 ETNQAQVDI-----LKKRTAEL-----EYNATALOKAKAVIDELEYN 635  
QY 618 QRP-----ACAPPELHSLRQLRERDVRINHLMDPEKRSQREBEKLLISAWNM 670  
DB 636 SRPVSDSNTSVQAQFKMEKEENELQKQ---KVEKLEIELNTVQGFQEQENRLLTSASHQ 692  
QY 671 -----GNAQQRAGEERAPAHAFSELAQOR-----LATNSRCPLGLRLASL 711  
DB 693 VLNRSIDVMSMAHAGSE-----EPOTLLDTOKMSGALPWRSIASETRRELPTAMASI 746

RESULT 8  
S21801  
myosin heavy chain, neuronal [similarity] - rat  
N: Alternate names: myosin II  
N: Contains: myosin ATPase (EC 3.6.4.1)  
C: Species: Rattus norvegicus (Norway rat)  
C: Date: 31-Dec-1993 #sequence, revision 31-Dec-1993 #text\_change 19-Apr-2002  
C: Accession: S21801; P00013; S18134

R: Sun, W.; Chantler, P.D.  
J. Mol. Biol. 224, 1185-1193, 1992  
A: Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain  
A: Reference number: S21801; MUID:92235856; PMID:1569576  
A: Accession: S21801  
A: Molecule type: mRNA  
A: Residues: 1-1999 <SUN>  
A: Cross-references: EMBL:X62659  
R: Sun, W.; Chantler, P.D.  
Biochem. Biophys. Res. Commun. 175, 244-249, 1991  
A: Title: A unique cellular myosin II exhibiting differential expression in the cerebellum  
A: Reference number: P00013; MUID:91151356; PMID:1998509  
A: Accession: P00013  
A: Molecule type: mRNA  
A: Residues: 1914-1998, 1' <SUN>  
A: Experimental source: brain  
C: Superfamily: myosin heavy chain; myosin motor domain homology  
C: Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide binding site; myosin motor domain homology <NMOT>  
F: 84-763/Domain: myosin motor domain homology motif A (P-loop)  
F: 174-181/Region: nucleotide-binding motif A (P-loop)  
F: 541-575/Region: actin binding #status predicted  
F: 653-675/Region: actin binding #status predicted  
F: 836-1999/Domain: coiled coil #status predicted <COI>  
F: 836-1276/Region: S2  
F: 1277-1999/Region: light meromyosin  
F: 125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted  
F: 180/Binding site: ATP (Lys) #status predicted  
F: 693,703/Active site: Cys #status predicted  
F: 1916/Binding site: phosphate (Ser) (covalent) #status predicted  
F: 1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 8.9%; Score 322; DB 1; Length 1999;  
Best Local Similarity 25.0%; Pred. No. 1.2e-06;  
Matches 159; Conservative 118; Mismatches 209; Indels 150; Gaps 31;

QY 109 AELGL-LQVL---GCAISCEK--ODHIQRIWT---LEESVQHVVMETAEOLMTK 156  
DB 970 AKLLKEEIIEDQNCLEKLEKLEDEFTNLTTEEKSKSLAKLNKHEAMIT 1029  
QY 157 DTPDLSPTYGNFDSQSRYYFLSEAEDEG--ELQRCILDLERQLMLLSEKQSLAQE 214  
DB 1030 DLEERLRE-----EKQKELEKTRKLEGSDTSLDQIAELQAO---IAELKMLAKK 1080  
QY 215 NAGLRMRGPGEGT-PGLTAKKLLQLQLEQOENFRLESGREDRLRCACELREEV 273  
DB 1081 EBELOAALARVEEERAAQKNNALKKIRELSQSELOED---LESE-----ASRNK 1128  
QY 274 AELQHNQALTSIAQEAQALKDEMDELROSS---BRAGOLEATLTSCRRRLGELRELK 329  
DB 1129 AKQRKD-----LGEELKTELEDTSTRAQOELRKRQEVNILKKTLEE---EAKT 1181  
QY 330 QYRQLEERNAGHAERTQLEDRLRAGSLRAQLEAQRQVQELQORQOENAKKAEKWLFE 389  
DB 1182 HEAQIQEMRQKHSQAVEELAEQLEQTKRVANLEKAK---QTLNERGELANEVKVLQ 1238  
QY 390 CRNLEKYESVYKKEKRLAERLAEERLREANE-ELRCALQPRGLTQADPSLDPTS---TPV 445  
DB 1239 GRDSEHKRKKVEAQELQV-----KFNEGERRVTEADK-VTKLOVELDNNVTGLLSOS 1291  
QY 446 DNLAAEI-----LPALRETLRLQLENK-----RLCROEAADRERQEEQLRHLEDA 492  
DB 1292 DSKSKLTDFSALESQLODTQELLEENRQKLSLTKLQVEDEKNSFREOLE---EEE 1348  
QY 493 NRAHGLETOHRLNQOQLSELRAQVEDLQKALQEQGGKTEDAISILLKLEHQLK---550  
DB 1349 EEAHNL-----KQIATLHAQVADMKKMKEDSVGCLTAEEV--KRKLOKDEGLSQ 1399  
QY 551 -HE-----ADLE-----LQKREYIELEPPTDSTAR 577  
DB 1400 RHEEKVAAYDKLEKTRQLQOQLDLDLVDLHORQSACNLEKQKQKFDQLAEITKSAK 1459  
QY 578 RIELQHNKKDADLRAEMEYRRYVVDKARMVMOTMEPKQRPAAAGAPPELHSLRLQRE 637

|    |     |                                                         |     |
|----|-----|---------------------------------------------------------|-----|
| QY | 476 | AADRERO-----EELQRHLEDANRARGLETOHRLNNOOLSELRAQVEDLQ----- | 521 |
|    |     |                                                         |     |
| Ds | 706 | --REORYLEKELOQEORLEKELOREDKRRQVRERKYLEEELQOEEDLRQEK     | 762 |
|    |     |                                                         |     |
| QY | 522 | KALQEOGGKTEDAISILKKLEHLQ-----KLHEADLELOR                | 559 |
|    |     |                                                         |     |
| Ds | 763 | QLNRDEBKRYLEKVELORE-EQLQREKRQREERYEEELLHRKEOOLQR        | 821 |
|    |     |                                                         |     |
| QY | 560 | KREYIELEPTDSTARRIEE-----LOHNLOKKADLRAMEERYVVKAR-----    | 608 |
|    |     |                                                         |     |
| Ds | 822 | ----EBCKRRRQELEROLEEELQRLDRSKROFRDDQHONEVNSRVYSKHRENKES | 876 |
|    |     |                                                         |     |
| QY | 609 | -----MYQMTEP-----KORPAAGAPPELHSLRTQIRE---               | 637 |
|    |     |                                                         |     |
| Ds | 877 | RQLDSDSVRESQPQDLPLQDOBEKREREQWRSKRDQSOPFABOLLEREQQEKTER | 936 |
|    |     |                                                         |     |
| QY | 638 | -----RDVIRHLEMDFEKSRSOREQEEKLLISAWYNMGMALOORAGEER       | 682 |
|    |     |                                                         |     |
| Ds | 937 | RDKFREEOQLKGQREKRYILEED---RKPREEQQL-----RRURERQOULRQER  | 985 |
|    |     |                                                         |     |
| QY | 683 | APAHAQSFLAQR                                            | 695 |
|    |     |                                                         |     |
| Ds | 986 | DRKFEELSQR                                              | 998 |
|    |     |                                                         |     |

RESULT 10  
A61231  
myosin heavy chain nonmuscle form A - human  
N;Alternate names: cellular myosin heavy chain; myosin type 9; NMHC-A  
N;Contains: myosin Arpase (EC 3.6.4.1)  
C;Species: Homo sapiens (man)  
C;Date: 12-May-1994 #sequence\_revision 14-Jul-1994 #text\_change 19-Apr-2002  
C;Accession: A61231; M3: A876; I52562; I61692  
R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein, R.S.  
Circ. Res. 69, 530-539, 1991  
A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on diff  
A;Reference number: A61231; MUID:91316803; PMID:1860190  
A;Accession: A61231  
A:Molecule type: mRNA  
A:Residues: 1-715 <SIM>  
A:Cross-references: GB:M6180; NID:g189029; PIDN:AAA61765.1; PID:g189030  
Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990  
R;Saez, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.  
A;Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through alt  
A;Reference number: A34876; MUID:90138958; PMID:1967836  
A;Accession: A34876  
A:Molecule type: mRNA  
A:Residues: 715-1961 <SAE>  
A:Cross-references: GB:M31013; NID:g189035; PIDN:AAA36349.1; PID:g189036  
R;Toothaker, L.E.; Gonzales, D.A.; Tung, N.; Lemons, R.S.; Le Beau, M.M.; Arnaut, M.P.  
Blood 78, 1826-1833, 1991  
A;Title: Cellular myosin heavy chain in human leukocytes: isolation of 5' cDNA clones  
A;Reference number: I52562; MUID:92003925; PMID:1912569  
A;Accession: I52562  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-52, 'EAI', 56-659, 'T', 661-868, 'T', 870-930, 'C', 932-1239, 'KG', 1242-1337 <RE  
A:Cross-references: GB:M81105; NID:g188988; PIDN:AAA5988.1; PID:g553596  
R;Bement, W.M.; Hasson, T.; Wirth, J.A.; Cheney, R.E.; Mooseker, M.S.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994  
A;Title: Identification and overlapping expression of multiple unconventional myosin  
A;Reference number: A55758; MUID:94294418; PMID:8022818  
A;Accession: I61692  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 182-218 <BEN>  
A:Cross-references: GB:L29141; NID:g457249; PIDN:AAA20904.1; PID:g531134  
C;Genetics:  
A;Gene: GDB:MYH9  
A:Cross-references: GDB:I20216; OMIM:160775  
A;Map position: 22q12.3-22q13.1  
C;Superfamily: myosin heavy chain; myosin motor domain homology  
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleoside

F:84-764/Domain: myosin motor domain homology <MMOT>  
F:174-181/Region: nucleotide-binding motif A (p-loop)  
F:552-565/Region: actin binding #status predicted  
F:626-640/Region: actin binding #status predicted  
F:837-1938/Domain: coiled coil #status predicted <COI>  
F:837-1277/Domain: S2 #status predicted <DS2>  
F:1278-1961/Domain: light meromyosin #status predicted <LMW>  
F:1939-1961/Domain: carboxyl-terminal <CBT>  
F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
F:180/Binding site: ATP (Lys) #status predicted  
F:694,704/Active site: Cys #status predicted

Query Match 8.7% Score 316; DB 1; Length 1961;  
Best Local Similarity 24.3%; Pred. No. 2.2e-06;  
Matches 153; Conservative 115; Mismatches 268; Indels 94; Gaps 24;

QY 109 AELGKL-LQLVL---GCAISCEKK-QDHQIRIMT-----LEESVQHVVMYAEIQELMTK 156  
DB 971 AKLKLLEEQIILEDQNCLEKLEEDRIAEFTNLTPEEKSLSLAKLNKHEAMIT 1030  
QY 157 DTPDSLPETYGNFDSQSRYYFLSEAEEDG-ELQQRCLDLERQLMLLSEKQSLAQE 214  
DB 1031 DLEERLRE-----EKQOELEKTRKLEGGDSTDLSDQIAELQAO---IAELKMLAKK 1081  
QY 215 NAGLRMRGRPEEGT-PGLTAKKLLILLASQLEQNEPRLESGRDELRCAELREV 273  
DB 1082 EEFQALARVEEAAQKNMALKIRELESQISELED---LESER-----ASRNK 1129  
QY 274 AELQHNQALTSQAQAQAKDMDELROSS---BRAGOLEATLTSCRRRLGELREL 329  
DB 1130 AEQKQD-----LGELEALKTELEDTLSTAAQOELRSAREQEVNLTLEE--EAKT 1182  
QY 330 QVRLERAGNAGHARTROLELRRAGSLRAQLEAQRQVQELQGORQOEAMAEKWLFE 389  
DB 1183 HEAQIQEMQKHQSAVEELAEQLEQTKRYKANLEKAK---QTLNENGERELANEVVLQ 1239  
QY 390 CRNLEKYESVTKERLLAERDSLRANEEL--RCAQLQPRGLTQADPSLDTSTPDVN 447  
DB 1240 GRDSEHKRKKVEAQLQLOVQKFNSEGRVETLADKYTKLOVE-LDNVTGLLSQSDSK 1298  
QY 448 LAAEI--LPAELRETLLRQLENNK-----RLCROEADREQOELORHLEDANRARG 498  
DB 1299 LTKDFSALESQLODTQELLQENRQKLSLTKLQVDEKNSFREQL-----EEEEKAHN 1355  
QY 499 LEPOHRLNOOQSELRAQVEDLQALQEOGGKTEDAISILLKRLKEHLQKLEADLEQ 558  
DB 1356 LE-----KQIATLHAQVADMKKMEDSVGCLTAEEV--KKLQKDLGSLQSREEKV 1406  
QY 559 RKREYIEEPPDSDSTARRIELOH-----NLOKD-----ADLRAMEERYRYVD 605  
DB 1407 AAYDKLEKTKRLQOQELDLVLDHQROSACNLEKKQKQKFDQLLAEKTIKAYAEED 1466  
QY 606 KARVMOTPEKORPAAGA-----PPELHSLRTOLEDRDVRIRHLEMDFEKRSRQREQ 658  
DB 1467 RAEAREKETKALSARALEEAMEQAELEKLNKOFRTMEDLMSSKDQVGSVHELEK 1526  
QY 659 EEKLLISANTNMGMALQORAGEERAPAHQ 688  
DB 1527 SKRALEQOQVEEMTKQLEELDEQLQATEDAK 1556

RESULT 11  
B35815  
myosin heavy chain 2, muscle - fruit fly (Drosophila melanogaster) (fragment)  
C:Species: Drosophila melanogaster  
C:Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 20-Jun-2000  
C:Accession: B35815  
R:Collier, V.L.; Kronert, W.A.; O'Donnell, P.T.; Edwards, K.A.; Bernstein, S.I.  
Genes Dev. 4, 885-895, 1990  
A:Title: Alternative myosin hinge regions are utilized in a tissue-specific fashion that  
A:Reference number: A35815; WUID:90346288; PMID:2116987  
A:Accession: B35815  
A>Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-1201 <COL>  
A:Cross-references: EMBL:X53155; NID:g8219; PIDN:CAA37311.1; PID:g2546939  
A>Note: the authors translated the codon TGC for residue 649 as Lys, and AAG for resi  
C:Genetics:  
A:Gene: FlyBase:Mhc  
A:Cross-references: FlyBase:FBgn0002741  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: ATP

Query Match 8.7% Score 315.5; DB 2; Length 1201;  
Best Local Similarity 21.3%; Pred. No. 1.4e-06;  
Matches 178; Conservative 144; Mismatches 271; Indels 241; Gaps 34;

QY 66 KLVSNLKMVLRSLVEYS---QDVLAHPVSEHLDPVSLIGESFSDPAELGKLQLVLGCA 122  
DB 403 KKEAEELSKLRRLDEEANIQHSTLAN-LRKKNDAVAEMAEQVD--QLNKLK-----A 453  
QY 123 ISEKKQDHTQRTMTLEESVQHVVMY-AIOELMTKTPDPSLSP-----ETYGNFDS 172  
DB 454 KAEKNEYTGQNLDRAGVDHITNEKAAQEKIAKQHTLNEVQSKLDENRTNLTNDFDA 513  
QY 173 QSRYYFLSEAEEDGELQORCLDLERQLMLS-----EEKSLAQENA-----216  
DB 514 SKKKL-----SIENSLRLQLEAESQVLSKIKISLTQLEDTKRLADEESRERATL 567  
QY 217-----GLRRMRGRPEEGT-PGLTAKKLLILLASQLEQLQEE-----251  
DB 568 LGKFRNLHLDNLREQV-EEEAEGRAD-----LQRLSKANAEAAQVWRSKYESDGV 618  
QY 252 --NFRLESGRDELRCAELEREVAELQHRNQALTSIAQEAQALKDMDELROSSER---306  
DB 619 ARSEEELEAKRKQARLAEAEETESL---NQKICIGLEKTKQRLSTEVEDLQLEVDRA 675  
QY 307-----AGOLEATLTSCR-----RRLGELRELRLQVRQ 333  
DB 676 IANAEEKQKAFKPIIGEWKLVYDDLAELDAASQKESYNTSEFLRLKGYEEGQOLEA 735  
QY 334 LEERNAGHARTROLELRRAG-----SLRAQLEAQRQVQ-----ELQK--ORQEE 379  
DB 736 VRENNKLADEVKLLDQIGEGGRNTHIEKARKRLAEKDEQLQAALEAEAELEENK 795  
QY 380 AMKAEKWLFCR-----NLEKYESVTKERLLAERDSLRANEELRCAQLQPRGL 431  
DB 796 VLRAQLELSQVROEIDRRIOKEEDEFENTRKNHQALDSMQASLEAEAKGKAEALRMKK 855  
QY 432 TOADPSLDPTSTPDVNLAAILPAELRETLLRQ-----LENKRLCRQEAADR---479  
DB 856 LEAD--INELEIALDH--ANKANAEAKNIRYCOQQLKQIQTALEEEQRRADDAEQLGI 911  
QY 480 -ERQ-----BELQRHLEDANRARGHLETOHRLNOOQSELRAQVEDLQKALQEOGK 530  
DB 912 SERRANALQNELESRTLLQADRGRAEQELADAHEQLNEVSAQNASISAA-----964  
QY 531 TEDAISILLKRLKEHLQKLEADL-----ELQRKREYI-----EELEPPPD--572  
DB 965 -----KKLESELEQLTH--SDLELLNEAKNSEKAKKAMVDAARLADLRAEQDHA 1014  
QY 573 -----SSTARIEELOHNLQKQDAD-LRAMEERYRYVDKARVMYQTMPEKORPAAGA 624  
DB 1015 QOEKRLKALEQQIKELQVLDLDEAEANALGKGGKAKIOKLEQVRRELENDGGEQRHADA 1074  
QY 625 PPELHSLRTOLEDRDVRIRHLEMDFEKRSRQREQEKL-----IS 665  
DB 1075 -----QKNLRSEKRVKELSFQSEDRKNHMRMDLVKQKIKTKYKQREAEIEA 1127  
QY 666 AWYNGMALQORAGEERAPAHQSFQAQRLA---TNSRRGLPGLRSLNLRPT 716  
DB 1128 A---LNLAKFRKAQOELEAEERADLAQIAISFRAGKGRAGSVGRGASAPRAT 1178

RESULT 12  
A45973

trichohyalin - human  
C:Species: Homo sapiens (man)  
C:Date: 03-May-1994 #sequence\_revision 01-Mar-1996 #text\_change 22-Jun-1999  
C:Accession: A45973  
R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.  
J. Biol. Chem. 268, 12164-12176, 1993  
A:Title: The structure of human trichohyalin. Potential multiple roles as a functional B  
ed (cross-linking) protein.  
A:Reference number: A45973; PMID:7685034  
A:Accession: A45973  
A:Molecule type: DNA  
A:Residues: 1-1898 <LEE>  
A:Cross-references: GB:L09190; NID:9292835; PIDN:AA65382.1; PID:9292836  
A:Note: authors translated the codon AGG for residue 1714 as Pro  
A:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh  
Covalent modifications to this protein include conversion of arginine to citrulline and  
C:Genetics:  
A:Gene: GDB:THH  
A:Cross-references: GDB:136223; OMTM:190370  
A:Map position: 1q21-1q21  
C:Superfamily: trichohyalin; calmodulin repeat homology  
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat  
F:49-81/Domain: calmodulin repeat homology <EF2>  
  
Query Match 8.7%; Score 315; DB 1; Length 1898;  
Best Local Similarity 24.5%; Pred. No. 2.4e-06;  
Matches 168; Conservative 110; Mismatches 269; Indels 140; Gaps 29;  
  
Qy 104 EFSDPAEIGK---LLQL-VLGCAISCEKKDHIQIMTLESVQ----HYVMEAIQELMTK 156  
Db 203 EPDDEQLRRLELURRGREKQKQORQDRVQEEKEKRRKRTVLRKEEKQ 262  
Qy 157 DTPD---SLSPETYNFDSQSRRY--YFSEAEAGDELLQRCLDLERQLMLLSSEKQL 211  
Db 263 EPQORLEQEE-----EQLAKLERQELRRERQEEQQOR-LRQQLRRQEEER 316  
Qy 212 AQENAGLRMRGPEGPGTGLTAKLLQLQLEQOLQENFRLSGRDELRCAELR 271  
Db 317 QOER--REQERRE-----QOERREQQLR---REQERREQQLRR 353  
Qy 272 EYAEQHRNQALSLAQE---AQALKDMDLRQSSERAGALEATITSCRRRLGELR--- 325  
Db 354 EQEE-ERREQQLRRQEEERREQQLRRQEEERREQQLRRQEEERREQQLRRQ 412  
Qy 326 ELRRVQRLQERNAGHAETRLQELRAGSLRAQLEAQRQVQVLOGQROEAMKAEK 385  
Db 413 QLRRQ-QLRRQQLRRQQLRRQQLRRQQLRRQEEERHE-QKHEQERREQLRKQ 470  
Qy 386 ---WLFECRLNEKYESVTKERKLLAERDSLREANEELRCAQLQPRGLTQADPSLDP 440  
Db 471 EERRDLW---KREETERHEQERRKQQLKRDQEEERBRWLKLEERREQ----- 519  
Qy 441 TSTPVDNLAAILPAELRETLRLQLENKRLCQEAADREQEEQLRHLEDANRAHGLE 500  
Db 520 -----ERREQQLRRQEEER--EQLKQEEERLQORLRSQQLRR--E 560  
Qy 501 TOHRLNQ--QQLSELRAQVEDLQALQEOGGTDAISILLKRLKEHLQKL---HEADL 555  
Db 561 QEBRLQOLLKREERLEQERREQLRKQEEERDQ---LLKREERROQLRKQEEERL 617  
Qy 556 ELORKREYIEELP-----PTSSSTARTEELQHNQ----- 587  
Db 618 EQLKREYERLQERREDRRLKREPEERREHLLKSEQERREQLRRQEEERREQ 677  
Qy 588 -KKADLRAMEERYRYVVKARVMQTMPEKORPA-----ACAPPELHSLRQLREDRYR 641  
Db 678 LKREERLEQLKREHEERREQLAEQEEQAEERIKSRIPKQWQOLESEADARQSK 737  
Qy 642 IRLHLEDFKRSRQEEKLLISA---WYNGMALQORAGEERA-----PAHAQSL 691  
Db 738 VLLEAPQAGAEAPQEEKRRRSELSQWEEERAHRRQEEERDFTWQWQAEKSR 797  
Qy 692 AQORLATNSRRGPGLRASLNLRPTDK 718

Db 798 GRQL---SARPPLEQRERQLRAEER 821

## RESULT 13

D35815

myosin heavy chain 4, muscle - fruit fly (Drosophila melanogaster) (fragment)  
C:Species: Drosophila melanogaster  
C:Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 13-Feb-1998  
A:Accession: D35815  
R:Collier, V.L.; Kronert, W.A.; O'Donnell, P.T.; Edwards, K.A.; Bernstein, S.I.  
Genes Dev. 4, 885-895, 1990  
A:Title: Alternative myosin hinge regions are utilized in a tissue-specific fashion t  
A:Reference number: A35815; PMID:90346288; PMID:2115987  
A:Accession: D35815  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1175 <COL>  
A:Cross-references: EMBL:X53155  
A:Note: the authors translated the codon TGC for residue 649 as Lys, and AAG for resi  
C:Genetics:  
A:Gene: FlyBase:Mhc  
A:Cross-references: FlyBase:FBgn0002741  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: ATP

Query Match 8.6%; Score 314.5; DB 2; Length 1175;  
Best Local Similarity 21.3%; Pred. No. 1.5e-06;  
Matches 176; Conservative 144; Mismatches 267; Indels 241; Gaps 34;

Qy 66 KUKVNLKMLVSLRSLVEYS---QDVLAHPVSEHLPDVLIGFSPDAEGLKLLQLVLGCA 122  
Db 403 KKREAEKSLRRDLSEANTQHESTLAN-LRKXNDVAEMAEQVD--QLNKLK-----A 453

Qy 123 ISCEKKQDHQIMTLESVQHVME-AIQELMTKTPDLSLP-----ETVGNFDS 172  
Db 454 KAEKNEYGYGLNDLRAGVDHITNEKAAQEKIAKQLOHTLNEVOSKLDTRNTLNDPFA 513

Qy 173 QGRVYFLSEAEEDQLQRCLDLERQLMLLS-----EKKOSLAQENA----- 216  
Db 514 SKKL-----SIENDDLRLQLEAEESQVLSKISKISLTQLEDTKRLADESRERATL 567

Qy 217 -----GLRMRGPEGTPGTAKKLLQLQLEQEE----- 251  
Db 568 LGKFRNLHDLNLRQV--EEAEQKAD-----LQRLSKANAEAVWRSKYESDGV 618

Qy 252 --NFRLESQREDELRCAELEREVAELQHRNQALTSLAQEAQALKDMDDELROSSER--- 306  
Db 619 ARSEELAEAKRQLQARLAEAEETIESL---NOKIGLEKTQRLSTEVEDLQLEVDRANA 675

Qy 307 -----AGLEATLTSR-----RRLGELRELRRQVRQ 333  
Db 676 IANAEEKCKAFDKIIGWKLVDDLAELDASQECRYSTELFLKGAEGEQOLEA 735

Qy 334 LEERNAGHAERTQLDELRAG-----SLRAQLEAQRQVQ-----ELQG---QRQEE 379  
Db 736 VRRKNKLADVKDLDDQIGEGGRNTHETAKARKLAEKDELQALAEAEAEAEENK 795

Qy 380 AKKAENKLFECR-----NLEEKYESVTKERKLLAERDSLREANEELRCAQLQPRGL 431  
Db 796 VLRAQLELSQVROEIDRRITQEEEFENTRKNHQRALDSMQASLEAEAKGAALRMKKK 855

Qy 432 TQADSLDPTSTPVONLAAEILPAELRETLRLQ-----LENKRLCROEAADR--- 479  
Db 856 LEAD--INELETALDH--ANKANAEAKNKKYQQQLKDQIQTALAEQEARDDARQOLGI 911

Qy 480 -ERQ-----EELQRLHLEDANRAHGLETHQRLNQQLSELRAQVEDLQKALQEQGK 930  
Db 912 SERRANALQNEESNTLLEQADRRQAEQELADAEHQLENEVSAQNASISAA----- 964

Qy 531 TEDATSIKRLKEHLQKLHEADL-----ELQKREYI-----EELPPTD--- 572  
Db 965 -----KRKLESELQTLH-SLDDELLNEAKNSEKAKAMVDAARLAEELRAEQDHA 1014

QY 573 -----SSTARIELOHNLQKAD-LRAMEERYRYVYDKARMVMQTMPEKQRPAGA 624  
 Db 1015 QTOEKLKALEQOIKELQVRLDEANALKGKKAQKLEQVRYLENELDGEORRHADA 1074  
 QY 625 PPELHSLRQLRERDVRIRHLEMDFKSRQSRQEEKLL-----IS 665  
 Db 1075 -----QNLKRSRVRVLSFSQSEDRKNHMRMODLVKLOQKIKTKYKQIEAEIA 1127  
 QY 666 ANYMGMAQORAGERAPAHQSFIAQORLA--TNSRRGPLGRLAS 710  
 Db 1128 A---LNLAKFRQAQOELEAEERADLAQAIKFRAGRAGSVGRGAS 1172  
 RESULT 14  
 T16416  
 hypothetical protein F52B10.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
 C:Accession: T16416  
 R:Wilcox, L.; Goodwin, E.B.; Szent-Gyorgyi, A.G.  
 submitted to the EMBL Data Library, December 1995  
 A:Description: The sequence of C. elegans cosmid F52B10.  
 A:Reference number: Z18510  
 A:Accession: T16416  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1956 <WIL>  
 A:Cross-references: EMBL:041990; NID:g1118154; PID:g1118155; PIDN:AAA83339.1; CESP:F52B10.1  
 C:Genetics:  
 A:Gene: CESP:F52B10.1  
 A:Introns: 44/2; 108/3; 161/1; 211/3; 242/3; 297/1; 345/1; 503/1; 876/3; 941/3; 1134/2;  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 F:81-763/Domain: myosin motor domain homology <MMO>  
 Query Match 8.6%; Score 314.5; DB 2; Length 1956;  
 Best Local Similarity 21.6%; Pred. No. 2.6e-06;  
 Matches 151; Conservative 136; Mismatches 247; Indels 165; Gaps 26;  
 QY 70 SNKMYRLSVEYSQDVLAHPVSEHLPDVSLIGFSDPAELGKLLQVLGCAISCEKKQ 129  
 Db 1043 SELEQHKRLLALED-----SKHLAE-----KMGKVEEL-----NNQLMKRD 1081  
 QY 130 DHTQRIWTL--EESVQVYVME-AIQELMTKDTDSL--SPETVGNFDSQ--SRRYFELS 181  
 Db 1082 EELQHLTRYDEESAVNTLMQKMDQMT--TIDELREDMETERNARKAEMTRREYVAQ 1139  
 QY 182 EEAEEDGELQORCLDLEROLM--LLSEKQSLAENAGLRMRGRPEG--EGTGPLTAKKL 238  
 Db 1140 LEKVGQVLDK--VDATMLQDLMSRKDEEVNATKRAIEQIHTMEGKIEQAKFSQV 1197  
 QY 239 LLQSOLEQOEENFRLESGREDELRCAELEREVAELQ-----HRNQAL 283  
 Db 1198 EELHDQIEQHKQORSQKQONQADQERADMAQIAALQASRADIDKRRKIHAEHLMEIQ 1257  
 QY 284 TSAQEAQALKDEMDLQSSERAGOL-----EATLTSRRRL--GELBELRQVR 332  
 Db 1258 ANLAESDEHKRTIUDLQSRDELHNRVREEEHAFANMORRLATAEQIOELNEQIQ 1317  
 QY 333 QLEBNAGHARTROLE-----DELRAGSLRAOLEAQ-----RRQVQLQ 374  
 Db 1318 EETRLKANTNARQLEDEKNALLDEKEEAGLRAHLEKIHARQAGARRKAESVN 1377  
 QY 375 QRQEEAMAEKWLFEENLKEYSVTKEKRLLAERDSLRANEELRCAOLQPRGLTQA 434  
 Db 1378 QOLEELAK--KNLRDVEHLQKQLESEVAKERTIQLSKKTIQ----- 1417  
 QY 435 DPSLDPTSTPDNLAAILPAELRETLRLQLENKRLC-----RQEAADRQ- 482  
 Db 1418 --ELEDSSMELENVSHRSDSEKQKRFESQMAERYAVQKALLDRDAMSQELRDRETRY 1475  
 QY 483 -----EELQHLEDANRAR-----HGLETHRLNQOOLSEL 514

Db 1476 LSLLEVDIMKEHLEESDRVRSLQOELQDSISNKDDFKNVHELEKAKRSLEAEIENDMR 1535  
 QY 515 AOVEDLQALQOEGGTEDA---ISILKRLKEHQLKHEADLELQKR---EYIEEL 567  
 Db 1536 VQWEELEDNLQ-----TAEDARLLEVTNALKSEDSRATSNKDVAAEKRRGLLKQIRDL 1591  
 QY 568 EPTDS-----STARRIELOHNLQKADLRAMEERYRYVYDKARMVMQTMPEKQ 618  
 Db 1592 ENELENEKRGKSGAVSHRKKIENQIGLEQOOLEVANRLKEEYVKNQKQIIEKQIIEC 1651  
 QY 619 RPAAGAPPELHSLRTOIRERDVRIRHLEMDFKSRQSRQ 657  
 Db 1652 EEARQAKEDIAAL---LREADKFRFAVEAREQLEANE 1687  
 RESULT 15  
 A40997  
 myosin heavy chain, striated adductor muscle - scallop (Aequipecten irradians)  
 N:Contains: myosin ATPase (EC 3.6.4.1)  
 C:Species: Aequipecten irradians  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Apr-2002  
 C:Accession: A40997; S13557  
 R:Nyitrai, L.; Goodwin, E.B.; Szent-Gyorgyi, A.G.  
 J. Biol. Chem. 266, 18469-18476, 1991  
 A:Title: Complete primary structure of a scallop striated muscle myosin heavy chain.  
 A:Reference number: A40997; MUID:92011595; PMID:1917970  
 A:Accession: A40997  
 A:Molecule type: mRNA  
 A:Residues: 1-1938 <NY>  
 A:Cross-references: GB:X55714; NID:g5611; PIDN:CAA39247.1; PID:g5612  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 C:Keywords: actin binding; ATP; coiled coil; hydrolase; muscle contraction; nucleotid  
 F:86-763/Domain: myosin motor domain homology <MMO>  
 F:176-183/Region: nucleotide-binding motif A (P-loop)  
 F:547-586/Region: actin binding #status predicted  
 F:653-675/Region: actin binding #status predicted  
 F:836-1938/Domain: actin binding #status predicted  
 F:836-1276/Region: S2  
 F:1277-1938/Region: light meromyosin  
 F:182/Binding site: ATP (Lys) #status predicted  
 F:693.703/Active site: Cys #status predicted  
 Query Match 8.6%; Score 312; DB 1; Length 1938;  
 Best Local Similarity 20.1%; Pred. No. 3.3e-06;  
 Matches 168; Conservative 153; Mismatches 287; Indels 228; Gaps 30;  
 QY 29 QDLSSGLAVYVLNQIDPSWF---NEAWLOGISEDGPNNMKL--KVSNLKMYRLSVEYS 83  
 Db 802 QDQRIGLS---VIQRIKRWLVLRNQW-----WKLYSKVKPLLSIARQEEEMK 847  
 QY 84 QDVLAHPVSEHLPDVSLIGFSDPAELGKLLQVLGCAISCEKKODHRIQIMTLEESVQ 143  
 Db 848 EQLKQMDKKEDLAKTERIKKEEQNVTL-----EOKNDLFLOQLTLEDMSG 896  
 QY 144 HVVMEATQELMTKDTPDLSLPETVGNFDSQSRVYFLSEAEESGDELQOQCLDLERQLML 203  
 Db 897 D-QEERVEKLMQK-----ADFESQIK-----ELEERLDEEADAAD 932  
 QY 204 LSEKQSLAENAGLRMRGRPE-----GEGTGPLTAKKLLQLQSOLEQ----- 247  
 Db 933 LEGIKKKMEADNANLKKIDGDLNTLQAEQDAKHDNQISTLQGEISQODEHIGLNKE 992  
 QY 248 ---LQENFLESGREDELRUC-----AELEREVAELQHR-----NQALTS 286  
 Db 993 KKALEEANKKTSLSQAEEDKCNHNLKAKLEQALDELEDNLEREKKVRGDVEKAKRV 1052  
 QY 287 AQAQALKQDEMDL---ROSSERAGOLEATLTSRRRLGE---LRELRRQVROLEERN 338  
 Db 1053 EQDLKTSQENVEDLVRVRELENVRNRKEAEISLSLKLEDEQNLVSQLOKRIKELQ--- 1109  
 QY 339 AGHAETRQLEDELRLAGSLRAQLEAQ---RQVQLOQORQEEA---MKA 383  
 Db 1110 ----ARIEELESELEARNARAKVQKQRAELNRELEL-GERLDEAGGATSQIELNKKR 1164

QY 384 EKWLFEC-RNLEE-----KYESVTKEKERLLAERDSLREA 417  
Db | : : : : :  
1165 EAELLKIRDLLEASLQHPAQISALUKKHQDAENEMADQVDOLQVKSKLEKDKDKKRE 1224  
QY 418 NEELR-----CAOLOPRGTLQADPSLDPTSTPVDNLAAEILPAELRETLRLQL 466  
Db | : : : : :  
1225 MDLLESQTHNMKNKGCSE---KVNKQFESOMDLNARLEDSQRSI--NELQSKSRLQA 1279  
QY 467 ENKRLCRO--EAADR-----ERQEELQRLHLEDANRHCLETOHRLNQOOLSE 512  
Db | : : : : :  
1280 ENSDLTRQLEDAEHRVSVLSKESQLSQLEDARKSLEETRARSKLQNEVRNMHADMDA 1339  
QY 513 LRAQVEDLQKA---LQEOGKGTEDAISILLKRLKLEHLQKLEADLELQKREYIEELEP 569  
Db | : : : : :  
1340 IREQLEEEQESDVQORQLSKANNELQWRSKFESEGANRTEELEDOKKRLGKLSEAEQ 1399  
QY 570 PTDSSTAR-----RIEELQHNLOKDDADLRAMEERYR-----RYVDKARM 609  
Db | : : : : :  
1400 TTEAANAKCSALEKAKSRLQOELEDMSEIYDRANASVNMQKQAFDKTTAEWQAKVNS 1459  
QY 610 VMOQTHPEKORPAAGAPPELHSLRTQLRERDVIRHLEM----- 647  
Db | : : : : :  
1460 LQSELENSQKESRGYSAEYRIKASIEEYQDSIGALRRKNKNLADEIHDLTDLQSEGGRS 1519  
QY 648 --DFEKSRSOROEKLLISAWYNGMALQQRAGEERAPAHQAQSLAQOORLATNSR 701  
Db | : : : : :  
1520 THELDKARRRLEMEKEELOQALEEAEGALEQ---EAKVMRAQLEIATVTRNEIDKR 1572

Search completed: March 4, 2003, 15:04:48  
Job time : 49.3446 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: March 4, 2003, 14:51:02 ; Search time 17.2109 Seconds  
(without alignments)  
1732.709 Million cell updates/sec

Title: US-09-763-902B-5  
Perfect score: 3638  
Sequence: 1 MSVDKAEKLGSLTLWLTQFH.....SRGFLGRLASLNRPTDKH 719

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description          |
|------------|-------|-------------|--------|--------------|----------------------|
| 1          | 360   | 9.9         | 1407   | 1 TRHY_RABIT | P37709 oryctolagus   |
| 2          | 344   | 9.5         | 4684   | 1 PLE1_HUMAN | Q15149 ratius sapien |
| 3          | 332.5 | 9.1         | 4687   | 1 PLE1_RAT   | P30427 ratius norv   |
| 4          | 328   | 9.0         | 4473   | 1 PLE1_CRIGR | Q91155 cricetus      |
| 5          | 321   | 8.8         | 1976   | 1 MYHA_BOVIN | Q27991 bos taurus    |
| 6          | 320.5 | 8.8         | 1961   | 1 MYH9_RAT   | Q62812 ratius norv   |
| 7          | 319   | 8.8         | 1549   | 1 TRHY_SHEEP | P22793 ovis aries    |
| 8          | 315   | 8.7         | 1898   | 1 TRHY_HUMAN | Q07283 homo sapien   |
| 9          | 313.5 | 8.6         | 1960   | 1 MYH9_HUMAN | P35579 homo sapien   |
| 10         | 312   | 8.6         | 1938   | 1 MYH9_ABOIR | P24733 aequipecten   |
| 11         | 306.5 | 8.4         | 1976   | 1 MYHA_HUMAN | P35580 homo sapien   |
| 12         | 306   | 8.4         | 886    | 1 RA50_ARCFU | Q29230 archaeglob    |
| 13         | 305.5 | 8.4         | 1935   | 1 MYSS_CYPCA | Q90339 cyprinus ca   |
| 14         | 304.5 | 8.4         | 1756   | 1 PEPL_HUMAN | O60437 homo sapien   |
| 15         | 302   | 8.3         | 978    | 1 RA50_AQUAE | O67124 aquifex aeo   |
| 16         | 301.5 | 8.3         | 1102   | 1 MYSC_CHICK | P29616 gallus gall   |
| 17         | 300   | 8.2         | 2230   | 1 GOG4_HUMAN | Q13439 homo sapien   |
| 18         | 299   | 8.2         | 1972   | 1 MYHB_HUMAN | P35749 homo sapien   |
| 19         | 297.5 | 8.2         | 1959   | 1 MYH9_CHICK | P14105 gallus gall   |
| 20         | 297.5 | 8.2         | 1962   | 1 MYSA_DROME | P05661 drosophila    |
| 21         | 297   | 8.2         | 857    | 1 MYEB_GIALA | Q08014 giardia lam   |
| 22         | 296   | 8.1         | 880    | 1 MYSP_BROMA | Q01202 brugia mala   |
| 23         | 296   | 8.1         | 1938   | 1 MYHD_HUMAN | P35748 oryctolagus   |
| 24         | 296   | 8.1         | 1972   | 1 MYHB_RABIT | Q9UKX3 homo sapien   |
| 25         | 296   | 8.1         | 1976   | 1 MYHA_RAT   | P49454 homo sapien   |
| 26         | 296   | 8.1         | 3210   | 1 MYSP_HUMAN | Q02171 onchocerca    |
| 27         | 296   | 8.0         | 879    | 1 MYSP_ONCVO | P25386 saccharomyc   |
| 28         | 292   | 8.0         | 1790   | 1 US01_YEAST | P12847 ratius norv   |
| 29         | 292   | 8.0         | 1940   | 1 MYH3_RAT   | Q99323 drosophila    |
| 30         | 292   | 8.0         | 2017   | 1 MYSN_DROME | P05659 acanthamoeb   |
| 31         | 290.5 | 8.0         | 1509   | 1 MYSN_ACACA | Q42184 gallus gall   |
| 32         | 289.5 | 8.0         | 1433   | 1 REST_CHICK | Q42184 gallus gall   |
| 33         | 288   | 7.9         | 1938   | 1 MYSS_CHICK | P13538 gallus gall   |

|    |       |     |      |              |                    |
|----|-------|-----|------|--------------|--------------------|
| 34 | 288   | 7.9 | 2871 | 1 DESP_HUMAN | P15924 homo sapien |
| 35 | 287   | 7.9 | 1939 | 1 MYH1_HUMAN | P12882 homo sapien |
| 36 | 287   | 7.9 | 1947 | 1 MYSC_CAEEL | P12845 caenorhabdi |
| 37 | 286.5 | 7.9 | 880  | 1 RA50_PIRAB | Q9UC28 pyrococcus  |
| 38 | 286   | 7.9 | 882  | 1 MYSP_CAEEL | P10567 caenorhabdi |
| 39 | 285.5 | 7.8 | 848  | 1 MYSP_DIRIM | P13392 dirofilaria |
| 40 | 285   | 7.8 | 1940 | 1 MYH3_CHICK | P02565 gallus gall |
| 41 | 285   | 7.8 | 1978 | 1 MYHB_CHICK | P10587 gallus gall |
| 42 | 284   | 7.8 | 1935 | 1 MYH7_HUMAN | P12883 homo sapien |
| 43 | 283   | 7.8 | 1427 | 1 REST_HUMAN | P30622 homo sapien |
| 44 | 282.5 | 7.8 | 1755 | 1 PEPL_MOUSE | Q9R269 mus musculu |
| 45 | 282.5 | 7.8 | 1935 | 1 MYH7_PIG   | P79293 sus scrofa  |

ALIGNMENTS

RESULT 1  
TRHY\_RABIT  
ID TRHY\_RABIT STANDARD; PRT; 1407 AA.  
AC P37709;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Trichohyalin.  
GN THH.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fietz M.J., Rogers G.E.;  
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES  
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER  
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN  
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN  
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL  
CC DIFFERENTIATION.  
CC -!- SUBUNIT: HOMODIMER (PROBABLE).  
CC -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS  
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN  
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF  
CC THE EPIDERMIS.  
CC -!- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND  
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST  
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS  
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRAINED  
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS  
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.  
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN  
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG  
CC DIFFERENT SPECIES.  
CC -!- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE  
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.  
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100  
CC FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
CC EMBL: Z19092; CAA79519.1;



DR PIR: S28589; S28589.  
 DR HSP: P02633; 41CB.  
 DR InterPro: IPR001751; CABP\_S100.  
 DR InterPro: IPR002048; EF-hand.  
 DR Pfam: PF00036; efhand; 1.  
 DR Pfam: PF01023; S\_100; 1.  
 DR ProDom: PD003407; CABP\_S100; 1.  
 DR ProSite: PS00018; EF\_HAND; 1.  
 DR ProSite: PS00303; S100\_CABP; 1.  
 KW Keratinization; Repeat; Calcium-binding.  
 FT DOMAIN 1 91  
 FT CA\_BIND 22 33  
 FT CA\_BIND 62 73  
 SQ SEQUENCE 1407 AA; 183781 MW; AE17D2A15F12B7F CRC64;  
 Query Match 9.9%; Score 360; DB 1; Length 1407;  
 Best Local Similarity 25.5%; Pred. No. 8.1e-09;  
 Matches 164; Conservative 110; Mismatches 262; Indels 108; Gaps 21;

QY 129 QDHQIRMTLESVHVHVAIQELMTKDTSPDTSPTGYGDFSDSRYYFLSEAE--E 186  
 DB 168 RKDEORLQRELEARRAEELQRRKGRDAEFEEELQRRREQLKRELEEEQQRRE 227  
 QY 167 GDELQRCGLDLERQLMLLS-----EKKSLAQENAGLRERMGPEE--GTFGLTAKK 237  
 DB 228 REQHERALQEEELQRRWRREPQEQQLRRELEIREERQLEERQLEERQLEERQ 287  
 QY 238 LL-----LQSQLEQLEQENFLESGR-----EDERLRCALEREVAELQHRNQL 283  
 DB 288 RLEOEERREQLRRELEIREERQLEERQLEERQLEERQLEERQLEERQLEERQ 347  
 QY 284 TSLAQEAQALDEWDE-----LRQSSERAGLEATLTSCRRRLGELRELRRQ 330  
 DB 348 EOEERRQLLEAEVREQAREGESLTRWQOLESEAGARQSVYRPRQEE-----QS 402  
 QY 331 VRQLEERNAGHAETROLEDRLRAGSLRA-----OLEAQRRQVQ 370  
 DB 403 LRQDQERRQ--RQERERELEEQARQOQWQAESEERRRQRLSARPSLBERQLRAEEERQ 461  
 QY 371 ELQQRQEEAMKAKWLFECRNLEEKVSVYKERRLAERDSLREANEELRCAQLQPRG 430  
 DB 462 E-QRFREEEQRER-RQELQLEEEQLRRAQAQLOEEDSFQEDRRRRRQEQRRPG 519  
 QY 431 LTQADPSLDPTSTVDNLAAILPAELRETLRLQLENKRLCROEADRDREQE--LQRH 488  
 DB 520 QTRWQLQEEAQRRTLYAK--PCQEQ--LREEELQREKRQEREREYREBEKLQRE 575  
 QY 489 LEDANRARGHLETOHRLNQOQLSELRAQVEDLQKALQGGKTEDAISILK-----RKL 543  
 DB 576 -EDKRRRQERQYR-----ELEELQEEQLDRKLEEEQLQREERLRRQERERKL 630  
 QY 544 EEHLQKHEADLELQKREYIELEPTDSTARRIEBELQHNLOKQKADLRAMEERYRY 603  
 DB 631 REEBQLRQEEQLRQERE-----RKLREEQLRREEQELR--QERERKL 674  
 QY 604 VDKAMVWQWTEPKORPAAGAPPELHSLRTOLRDVRIHLEMDFKSRQSRQEKLL 663  
 DB 675 REEBQLRQEEERLRRQERA-----RKLREEQLRQEEQLRQERERKRLREBEQL 726  
 QY 664 ISAWYNGMALQQRAGERAPAPHAQSLAQORLATNSRRGPLGR 707  
 DB 727 LRR--EQQLRQERDRKLEEEQLQSEERLRRRQEQQLRR 768

RESULT 2  
 ID PLE1\_HUMAN STANDARD: PRT; 4684 AA.  
 AC Q15149; Q16640; Q15148;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1).  
 SN PLEC1

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Placenta;  
 RX MEDLINE=96210632; PubMed=8633055;  
 RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;  
 RT "Human plectin: organization of the gene, sequence analysis, and  
 RT chromosome localization (8q24).";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.  
 RX MEDLINE=96312447; PubMed=8698233;  
 RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,  
 RA Bullrich F., Burgeson R.E., Amato S., Hudson D.L., Owaribe K.,  
 RA McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,  
 RA Utito J.;  
 RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:  
 RT cDNA cloning and genomic organization.";  
 RL Genes Dev. 10:1724-1735(1996).  
 RN [3]  
 RP VARIANT MD-EBS 1003-GLN--ALA-1005 DEL.  
 RX MEDLINE=97049959; PubMed=8894687;  
 RA Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,  
 RA Hachisuka H., Nishikawa T., McLean W.H.I., Utito J.;  
 RT "Homozygous deletion mutations in the plectin gene (PLEC1) in patients  
 RT with epidermolysis bullosa simplex associated with late-onset  
 RT muscular dystrophy.";  
 RL Hum. Mol. Genet. 5:1539-1546(1996).  
 RN [4]  
 RP VARIANT MD-EBS LEU-429 INS.  
 RX MEDLINE=21090821; PubMed=11159198;  
 RA Bauer J.W., Rouan F., Kofler B., Reznicek G.A., Kornacker I.,  
 RA Muss W., Hanthner R., Klausberger A., Huber A., Pohla-Gubo G.,  
 RA Wiche G., Utito J., Hanthner H.;  
 RT "A compound heterozygous one amino-acid insertion/nonsense mutation in  
 RT the plectin gene causes epidermolysis bullosa simplex with plectin  
 RT deficiency.";  
 RL Am. J. Pathol. 158:617-625(2001).  
 CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND  
 CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSESOMES OR  
 CC HEMIDESMOSESOMES. COULD ALSO BIND MUSCLE PROTEINS SUCH AS ACTIN TO  
 CC MEMBRANE COMPLEXES IN MUSCLE. MAY BE INVOLVED NOT ONLY IN THE  
 CC CROSSLINKING AND STABILIZATION OF CYTOSKELETAL INTERMEDIATE  
 CC FILAMENTS NETWORK, BUT ALSO IN THE REGULATION OF THEIR DYNAMICS.  
 CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN  
 CC MUSCLE, HEART, PLACENTA AND SPINAL CORD.  
 CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH  
 CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-  
 CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.  
 CC -1- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM  
 CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).  
 CC -1- DISEASE: DEFECTS IN PLEC1 ARE THE CAUSE OF EPIDERMOLYSIS BULLOSA  
 CC WITH MUSCULAR DYSTROPHY (MD-EBS OR EB-MD), AN AUTOSOMAL RECESSIVE  
 CC DISORDER CHARACTERIZED BY EPIDERMAL BLISTER FORMATION AT THE LEVEL  
 CC OF THE HEMIDESMOSE AND ASSOCIATED LATE-ONSET MUSCULAR DYSTROPHY.  
 CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 33 CALPONIN-HOMOLOGY (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE PLEKIN OR CYTOLINKER FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)



```

1566 SEAEIQAKARQAQAARSRLRIEETRVRLQLEATERGGAEGELQALRAAEAEAAQ 1625
341 -----HAER-TROLEDELR-----AGSLQAQLEAQRQVQELQ--QRQEPAMKA 383
1626 KROAQEAERLRQVQDESQRKROAEVSLASRVKAEAAAREKORALQALELRLQAEAA 1685
384 EKWLFE-----CRNLSEKVESYTKERLLAERDLSREANEELRCALQPRGLTQADPSL 438
1686 ERWLQCAEYERARQVQALETAQSAE---AELOSKRASFAE-KTAQLE-RSLQEEHVAV 1740
439 DPTSTPDVNLAEILPAE-----LRETL-LRLQ-----LENKRLCRQEA- 476
1741 AQLREAEARRAQQAQAEARAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 1800
477 -----ADRE-----ROEEL-QRHLEDANARHGLETOHRLNQOQSELRA--- 515
1801 KQKEAEERARRGKAEQAVRQELAEQELKQRLAEGTAQORLAEEAEQELRLRAETE 1860
516 QVEDLOKALQEOGQKTEDAISILKKRLEHLOKLHEADLELQKREYIEEPPDTSST 575
1861 QGQOQLLEELAR-----LQREAAATOKQOELEAEALAKVRAEMEVL---ASK 1908
576 ARRIEELQHNLOKQADLRAEMEERYRYVDKA-----BWMQTMPEKORPAA 622
1909 AKAEESRSTSEKSKORLEAEAGRFELAEAEARLALAEAEAKRQQLAEEDAAORAEA 1968
623 -----GAPPELSHRLTOLREDRVIRHLEMDFKSRQSEQEKLLISAWYNNGMAL 674
1969 ERVLAELKLAIGENTRLKT---EAEIALKEKEAENERLRLAEDE-----AF 2012
675 QQRAGERAPAH---AQSLAQOORLATNS 700
2013 QRRRLFEQAQHKRAIDIEERLAQLRKASDS 2041

RESULT 3
LEL_RAT
D PLEL_RAT STANDARD; PRT; 4687 AA.
C P30427; O08879; O08880; O08881;
T 01-APR-1993 (Rel. 25, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Plectin 1 (PLTN) (PCN).
S Rattus norvegicus (Rat).
X Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
X Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
X NCBI_TaxID=10116;
X [1]
X SEQUENCE FROM N.A. (ISOFORM 1).
X TISSUE=Glial tumor;
X MEDLINE=91268156; PubMed=2050743;
X Wiche G., Becker B., Luber K., Weitzer G., Castanon M.J.,
X Hauptmann R., Stratowa C., Stewart M.;
X "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
X chain with a three-domain structure based on a central alpha-helical
X coiled coil.";
X J. Cell Biol. 114:83-99(1991).
X [2]
X REVISIONS.
X TISSUE=Glial tumor;
X MEDLINE=96210632; PubMed=8633055;
X Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
X "Human plectin: organization of the gene, sequence analysis, and
X chromosome localization (8q24).";
X Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
X [3]
X PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4), AND TISSUE SPECIFICITY.
X TISSUE=Glial tumor;
X MEDLINE=97321050; PubMed=9177781;
X Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,
X Wiche G.;

```

```

"plectin transcript diversity: identification and tissue distribution
of variants with distinct first coding exons and rodless isoforms.";
Genomics 42:115-125(1997).
-!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSESOMES OR
HEMIDESMOSESOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
ALSO IN THE REGULATION OF THEIR DYNAMICS.
-!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
-!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE
PRODUCED BY ALTERNATIVE SPLICING.
-!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN
SKELETAL MUSCLE AND LOWEST IN THYMUS.
-!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
-!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
-!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
-!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
-!- SIMILARITY: CONTAINS 33 PLECTIN REPEATS.
-!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
-!- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
-!- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X59601; CAA42169.1; -
DR EMBL; U96274; AAC53209.1; -
DR EMBL; U96275; AAC53210.1; -
DR EMBL; U96276; AAC53211.1; -
DR PIR; A39638; A39638.
DR PIR; S21876; S21876.
DR HSP; Q01082; IBRK.
DR InterPro; IPR001589; Actbind_actnln.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR001101; Plectin_repeat.
DR InterPro; IPR005326; S10_plectin.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF03037; CH; 2.
DR Pfam; PF00681; Plectin; 21.
DR Pfam; PF03501; S10_plectin; 1.
DR SMART; SM00033; CH; 2.
DR SMART; SM00250; PLEC; 33.
DR SMART; SM00150; SPEC; 4.
DR PROSITE; PS00019; ACTININ_1; FALSE_NEG.
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE; PS00021; CH; 2.
KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
KW Phosphorylation; Alternative splicing.
FT DOMAIN 1 1473 GLOBULAR 1.
FT DOMAIN 1474 2758 CENTRAL FIBROUS ROD DOMAIN.
FT DOMAIN 2759 4687 GLOBULAR 2.
FT DOMAIN 181 406 ACTIN-BINDING.
FT DOMAIN 185 288 CH 1.
FT DOMAIN 301 403 CH 2.
FT REPEAT 648 722 SPECTRIN 1.
FT REPEAT 743 827 SPECTRIN 2.
FT REPEAT 840 933 SPECTRIN 3.
FT REPEAT 1318 1418 SPECTRIN 4.
FT DOMAIN 1472 1692 COILED COIL (POTENTIAL).
FT DOMAIN 1724 2760 COILED COIL (POTENTIAL).
FT REPEAT 2791 2828 PLECTIN 1.
FT REPEAT 2829 2866 PLECTIN 2.
FT REPEAT 2867 2904 PLECTIN 3.
FT REPEAT 2905 2942 PLECTIN 4.
FT REPEAT 2943 2980 PLECTIN 5.

```

```

FT REPEAT 2984 3018 PLECTIN 6.
FT REPEAT 3119 3156 PLECTIN 7.
FT REPEAT 3157 3194 PLECTIN 8.
FT REPEAT 3195 3232 PLECTIN 9.
FT REPEAT 3233 3270 PLECTIN 10.
FT REPEAT 3271 3308 PLECTIN 11.
FT REPEAT 3311 3346 PLECTIN 12.
FT REPEAT 3348 3385 PLECTIN 13.
FT REPEAT 3386 3423 PLECTIN 14.
FT REPEAT 3424 3461 PLECTIN 15.
FT REPEAT 3462 3500 PLECTIN 16.
FT REPEAT 3501 3538 PLECTIN 17.
FT REPEAT 3539 3576 PLECTIN 18.
FT REPEAT 3577 3614 PLECTIN 19.
FT REPEAT 3615 3652 PLECTIN 20.
FT REPEAT 3653 3690 PLECTIN 21.
FT REPEAT 3691 3728 PLECTIN 22.
FT REPEAT 3729 3766 PLECTIN 23.
FT REPEAT 3767 3804 PLECTIN 24.
FT REPEAT 3805 3842 PLECTIN 25.
FT REPEAT 3843 3880 PLECTIN 26.
FT REPEAT 3881 3918 PLECTIN 27.
FT REPEAT 3919 3956 PLECTIN 28.
FT REPEAT 3957 3994 PLECTIN 29.
FT REPEAT 3995 4032 PLECTIN 30.
FT REPEAT 4033 4070 PLECTIN 31.
FT REPEAT 4071 4108 PLECTIN 32.
FT REPEAT 4109 4146 PLECTIN 33.
FT REPEAT 4147 4184 PLECTIN 34.
FT REPEAT 4185 4222 PLECTIN 35.
FT REPEAT 4223 4260 PLECTIN 36.
FT REPEAT 4261 4298 PLECTIN 37.
FT REPEAT 4299 4336 PLECTIN 38.
FT REPEAT 4337 4374 PLECTIN 39.
FT REPEAT 4375 4412 PLECTIN 40.
FT REPEAT 4413 4450 PLECTIN 41.
FT REPEAT 4451 4488 PLECTIN 42.
FT REPEAT 4489 4526 PLECTIN 43.
FT REPEAT 4527 4564 PLECTIN 44.
FT REPEAT 4565 4602 PLECTIN 45.
FT REPEAT 4603 4640 PLECTIN 46.
FT REPEAT 4641 4678 PLECTIN 47.
FT REPEAT 4679 4716 PLECTIN 48.
FT REPEAT 4717 4754 PLECTIN 49.
FT REPEAT 4755 4792 PLECTIN 50.
FT REPEAT 4793 4830 PLECTIN 51.
FT REPEAT 4831 4868 PLECTIN 52.
FT REPEAT 4869 4906 PLECTIN 53.
FT REPEAT 4907 4944 PLECTIN 54.
FT REPEAT 4945 4982 PLECTIN 55.
FT REPEAT 4983 5020 PLECTIN 56.
FT REPEAT 5021 5058 PLECTIN 57.
FT REPEAT 5059 5096 PLECTIN 58.
FT REPEAT 5097 5134 PLECTIN 59.
FT REPEAT 5135 5172 PLECTIN 60.
FT REPEAT 5173 5210 PLECTIN 61.
FT REPEAT 5211 5248 PLECTIN 62.
FT REPEAT 5249 5286 PLECTIN 63.
FT REPEAT 5287 5324 PLECTIN 64.
FT REPEAT 5325 5362 PLECTIN 65.
FT REPEAT 5363 5400 PLECTIN 66.
FT REPEAT 5401 5438 PLECTIN 67.
FT REPEAT 5439 5476 PLECTIN 68.
FT REPEAT 5477 5514 PLECTIN 69.
FT REPEAT 5515 5552 PLECTIN 70.
FT REPEAT 5553 5590 PLECTIN 71.
FT REPEAT 5591 5628 PLECTIN 72.
FT REPEAT 5629 5666 PLECTIN 73.
FT REPEAT 5667 5704 PLECTIN 74.
FT REPEAT 5705 5742 PLECTIN 75.
FT REPEAT 5743 5780 PLECTIN 76.
FT REPEAT 5781 5818 PLECTIN 77.
FT REPEAT 5819 5856 PLECTIN 78.
FT REPEAT 5857 5894 PLECTIN 79.
FT REPEAT 5895 5932 PLECTIN 80.
FT REPEAT 5933 5970 PLECTIN 81.
FT REPEAT 5971 6008 PLECTIN 82.
FT REPEAT 6009 6046 PLECTIN 83.
FT REPEAT 6047 6084 PLECTIN 84.
FT REPEAT 6085 6122 PLECTIN 85.
FT REPEAT 6123 6160 PLECTIN 86.
FT REPEAT 6161 6198 PLECTIN 87.
FT REPEAT 6199 6236 PLECTIN 88.
FT REPEAT 6237 6274 PLECTIN 89.
FT REPEAT 6275 6312 PLECTIN 90.
FT REPEAT 6313 6350 PLECTIN 91.
FT REPEAT 6351 6388 PLECTIN 92.
FT REPEAT 6389 6426 PLECTIN 93.
FT REPEAT 6427 6464 PLECTIN 94.
FT REPEAT 6465 6502 PLECTIN 95.
FT REPEAT 6503 6540 PLECTIN 96.
FT REPEAT 6541 6578 PLECTIN 97.
FT REPEAT 6579 6616 PLECTIN 98.
FT REPEAT 6617 6654 PLECTIN 99.
FT REPEAT 6655 6692 PLECTIN 100.
FT REPEAT 6693 6730 PLECTIN 101.
FT REPEAT 6731 6768 PLECTIN 102.
FT REPEAT 6769 6806 PLECTIN 103.
FT REPEAT 6807 6844 PLECTIN 104.
FT REPEAT 6845 6882 PLECTIN 105.
FT REPEAT 6883 6920 PLECTIN 106.
FT REPEAT 6921 6958 PLECTIN 107.
FT REPEAT 6959 6996 PLECTIN 108.
FT REPEAT 6997 7034 PLECTIN 109.
FT REPEAT 7035 7072 PLECTIN 110.
FT REPEAT 7073 7110 PLECTIN 111.
FT REPEAT 7111 7148 PLECTIN 112.
FT REPEAT 7149 7186 PLECTIN 113.
FT REPEAT 7187 7224 PLECTIN 114.
FT REPEAT 7225 7262 PLECTIN 115.
FT REPEAT 7263 7300 PLECTIN 116.
FT REPEAT 7301 7338 PLECTIN 117.
FT REPEAT 7339 7376 PLECTIN 118.
FT REPEAT 7377 7414 PLECTIN 119.
FT REPEAT 7415 7452 PLECTIN 120.
FT REPEAT 7453 7490 PLECTIN 121.
FT REPEAT 7491 7528 PLECTIN 122.
FT REPEAT 7529 7566 PLECTIN 123.
FT REPEAT 7567 7604 PLECTIN 124.
FT REPEAT 7605 7642 PLECTIN 125.
FT REPEAT 7643 7680 PLECTIN 126.
FT REPEAT 7681 7718 PLECTIN 127.
FT REPEAT 7719 7756 PLECTIN 128.
FT REPEAT 7757 7794 PLECTIN 129.
FT REPEAT 7795 7832 PLECTIN 130.
FT REPEAT 7833 7870 PLECTIN 131.
FT REPEAT 7871 7908 PLECTIN 132.
FT REPEAT 7909 7946 PLECTIN 133.
FT REPEAT 7947 7984 PLECTIN 134.
FT REPEAT 7985 8022 PLECTIN 135.
FT REPEAT 8023 8060 PLECTIN 136.
FT REPEAT 8061 8098 PLECTIN 137.
FT REPEAT 8099 8136 PLECTIN 138.
FT REPEAT 8137 8174 PLECTIN 139.
FT REPEAT 8175 8212 PLECTIN 140.
FT REPEAT 8213 8250 PLECTIN 141.
FT REPEAT 8251 8288 PLECTIN 142.
FT REPEAT 8289 8326 PLECTIN 143.
FT REPEAT 8327 8364 PLECTIN 144.
FT REPEAT 8365 8402 PLECTIN 145.
FT REPEAT 8403 8440 PLECTIN 146.
FT REPEAT 8441 8478 PLECTIN 147.
FT REPEAT 8479 8516 PLECTIN 148.
FT REPEAT 8517 8554 PLECTIN 149.
FT REPEAT 8555 8592 PLECTIN 150.
FT REPEAT 8593 8630 PLECTIN 151.
FT REPEAT 8631 8668 PLECTIN 152.
FT REPEAT 8669 8706 PLECTIN 153.
FT REPEAT 8707 8744 PLECTIN 154.
FT REPEAT 8745 8782 PLECTIN 155.
FT REPEAT 8783 8820 PLECTIN 156.
FT REPEAT 8821 8858 PLECTIN 157.
FT REPEAT 8859 8896 PLECTIN 158.
FT REPEAT 8897 8934 PLECTIN 159.
FT REPEAT 8935 8972 PLECTIN 160.
FT REPEAT 8973 9010 PLECTIN 161.
FT REPEAT 9011 9048 PLECTIN 162.
FT REPEAT 9049 9086 PLECTIN 163.
FT REPEAT 9087 9124 PLECTIN 164.
FT REPEAT 9125 9162 PLECTIN 165.
FT REPEAT 9163 9200 PLECTIN 166.
FT REPEAT 9201 9238 PLECTIN 167.
FT REPEAT 9239 9276 PLECTIN 168.
FT REPEAT 9277 9314 PLECTIN 169.
FT REPEAT 9315 9352 PLECTIN 170.
FT REPEAT 9353 9390 PLECTIN 171.
FT REPEAT 9391 9428 PLECTIN 172.
FT REPEAT 9429 9466 PLECTIN 173.
FT REPEAT 9467 9504 PLECTIN 174.
FT REPEAT 9505 9542 PLECTIN 175.
FT REPEAT 9543 9580 PLECTIN 176.
FT REPEAT 9581 9618 PLECTIN 177.
FT REPEAT 9619 9656 PLECTIN 178.
FT REPEAT 9657 9694 PLECTIN 179.
FT REPEAT 9695 9732 PLECTIN 180.
FT REPEAT 9733 9770 PLECTIN 181.
FT REPEAT 9771 9808 PLECTIN 182.
FT REPEAT 9809 9846 PLECTIN 183.
FT REPEAT 9847 9884 PLECTIN 184.
FT REPEAT 9885 9922 PLECTIN 185.
FT REPEAT 9923 9960 PLECTIN 186.
FT REPEAT 9961 10000 PLECTIN 187.
FT REPEAT 10001 10040 PLECTIN 188.
FT REPEAT 10041 10080 PLECTIN 189.
FT REPEAT 10081 10120 PLECTIN 190.
FT REPEAT 10121 10160 PLECTIN 191.
FT REPEAT 10161 10200 PLECTIN 192.
FT REPEAT 10201 10240 PLECTIN 193.
FT REPEAT 10241 10280 PLECTIN 194.
FT REPEAT 10281 10320 PLECTIN 195.
FT REPEAT 10321 10360 PLECTIN 196.
FT REPEAT 10361 10400 PLECTIN 197.
FT REPEAT 10401 10440 PLECTIN 198.
FT REPEAT 10441 10480 PLECTIN 199.
FT REPEAT 10481 10520 PLECTIN 200.
FT REPEAT 10521 10560 PLECTIN 201.
FT REPEAT 10561 10600 PLECTIN 202.
FT REPEAT 10601 10640 PLECTIN 203.
FT REPEAT 10641 10680 PLECTIN 204.
FT REPEAT 10681 10720 PLECTIN 205.
FT REPEAT 10721 10760 PLECTIN 206.
FT REPEAT 10761 10800 PLECTIN 207.
FT REPEAT 10801 10840 PLECTIN 208.
FT REPEAT 10841 10880 PLECTIN 209.
FT REPEAT 10881 10920 PLECTIN 210.
FT REPEAT 10921 10960 PLECTIN 211.
FT REPEAT 10961 11000 PLECTIN 212.
FT REPEAT 11001 11040 PLECTIN 213.
FT REPEAT 11041 11080 PLECTIN 214.
FT REPEAT 11081 11120 PLECTIN 215.
FT REPEAT 11121 11160 PLECTIN 216.
FT REPEAT 11161 11200 PLECTIN 217.
FT REPEAT 11201 11240 PLECTIN 218.
FT REPEAT 11241 11280 PLECTIN 219.
FT REPEAT 11281 11320 PLECTIN 220.
FT REPEAT 11321 11360 PLECTIN 221.
FT REPEAT 11361 11400 PLECTIN 222.
FT REPEAT 11401 11440 PLECTIN 223.
FT REPEAT 11441 11480 PLECTIN 224.
FT REPEAT 11481 11520 PLECTIN 225.
FT REPEAT 11521 11560 PLECTIN 226.
FT REPEAT 11561 11600 PLECTIN 227.
FT REPEAT 11601 11640 PLECTIN 228.
FT REPEAT 11641 11680 PLECTIN 229.
FT REPEAT 11681 11720 PLECTIN 230.
FT REPEAT 11721 11760 PLECTIN 231.
FT REPEAT 11761 11800 PLECTIN 232.
FT REPEAT 11801 11840 PLECTIN 233.
FT REPEAT 11841 11880 PLECTIN 234.
FT REPEAT 11881 11920 PLECTIN 235.
FT REPEAT 11921 11960 PLECTIN 236.
FT REPEAT 11961 12000 PLECTIN 237.
FT REPEAT 12001 12040 PLECTIN 238.
FT REPEAT 12041 12080 PLECTIN 239.
FT REPEAT 12081 12120 PLECTIN 240.
FT REPEAT 12121 12160 PLECTIN 241.
FT REPEAT 12161 12200 PLECTIN 242.
FT REPEAT 12201 12240 PLECTIN 243.
FT REPEAT 12241 12280 PLECTIN 244.
FT REPEAT 12281 12320 PLECTIN 245.
FT REPEAT 12321 12360 PLECTIN 246.
FT REPEAT 12361 12400 PLECTIN 247.
FT REPEAT 12401 12440 PLECTIN 248.
FT REPEAT 12441 12480 PLECTIN 249.
FT REPEAT 12481 12520 PLECTIN 250.
FT REPEAT 12521 12560 PLECTIN 251.
FT REPEAT 12561 12600 PLECTIN 252.
FT REPEAT 12601 12640 PLECTIN 253.
FT REPEAT 12641 12680 PLECTIN 254.
FT REPEAT 12681 12720 PLECTIN 255.
FT REPEAT 12721 12760 PLECTIN 256.
FT REPEAT 12761 12800 PLECTIN 257.
FT REPEAT 12801 12840 PLECTIN 258.
FT REPEAT 12841 12880 PLECTIN 259.
FT REPEAT 12881 12920 PLECTIN 260.
FT REPEAT 12921 12960 PLECTIN 261.
FT REPEAT 12961 13000 PLECTIN 262.
FT REPEAT 13001 13040 PLECTIN 263.
FT REPEAT 13041 13080 PLECTIN 264.
FT REPEAT 13081 13120 PLECTIN 265.
FT REPEAT 13121 13160 PLECTIN 266.
FT REPEAT 13161 13200 PLECTIN 267.
FT REPEAT 13201 13240 PLECTIN 268.
FT REPEAT 13241 13280 PLECTIN 269.
FT REPEAT 13281 13320 PLECTIN 270.
FT REPEAT 13321 13360 PLECTIN 271.
FT REPEAT 13361 13400 PLECTIN 272.
FT REPEAT 13401 13440 PLECTIN 273.
FT REPEAT 13441 13480 PLECTIN 274.
FT REPEAT 13481 13520 PLECTIN 275.
FT REPEAT 13521 13560 PLECTIN 276.
FT REPEAT 13561 13600 PLECTIN 277.
FT REPEAT 13601 13640 PLECTIN 278.
FT REPEAT 13641 13680 PLECTIN 279.
FT REPEAT 13681 13720 PLECTIN 280.
FT REPEAT 13721 13760 PLECTIN 281.
FT REPEAT 13761 13800 PLECTIN 282.
FT REPEAT 13801 13840 PLECTIN 283.
FT REPEAT 13841 13880 PLECTIN 284.
FT REPEAT 13881 13920 PLECTIN 285.
FT REPEAT 13921 13960 PLECTIN 286.
FT REPEAT 13961 14000 PLECTIN 287.
FT REPEAT 14001 14040 PLECTIN 288.
FT REPEAT 14041 14080 PLECTIN 289.
FT REPEAT 14081 14120 PLECTIN 290.
FT REPEAT 14121 14160 PLECTIN 291.
FT REPEAT 14161 14200 PLECTIN 292.
FT REPEAT 14201 14240 PLECTIN 293.
FT REPEAT 14241 14280 PLECTIN 294.
FT REPEAT 14281 14320 PLECTIN 295.
FT REPEAT 14321 14360 PLECTIN 296.
FT REPEAT 14361 14400 PLECTIN 297.
FT REPEAT 14401 14440 PLECTIN 298.
FT REPEAT 14441 14480 PLECTIN 299.
FT REPEAT 14481 14520 PLECTIN 300.
FT REPEAT 14521 14560 PLECTIN 301.
FT REPEAT 14561 14600 PLECTIN 302.
FT REPEAT 14601 14640 PLECTIN 303.
FT REPEAT 14641 14680 PLECTIN 304.
FT REPEAT 14681 14720 PLECTIN 305.
FT REPEAT 14721 14760 PLECTIN 306.
FT REPEAT 14761 14800 PLECTIN 307.
FT REPEAT 14801 14840 PLECTIN 308.
FT REPEAT 14841 14880 PLECTIN 309.
FT REPEAT 14881 14920 PLECTIN 310.
FT REPEAT 14921 14960 PLECTIN 311.
FT REPEAT 14961 15000 PLECTIN 312.
FT REPEAT 15001 15040 PLECTIN 313.
FT REPEAT 15041 15080 PLECTIN 314.
FT REPEAT 15081 15120 PLECTIN 315.
FT REPEAT 15121 15160 PLECTIN 316.
FT REPEAT 15161 15200 PLECTIN 317.
FT REPEAT 15201 15240 PLECTIN 318.
FT REPEAT 15241 15280 PLECTIN 319.
FT REPEAT 15281 15320 PLECTIN 320.
FT REPEAT 15321 15360 PLECTIN 321.
FT REPEAT 15361 15400 PLECTIN 322.
FT REPEAT 15401 15440 PLECTIN 323.
FT REPEAT 15441 15480 PLECTIN 324.
FT REPEAT 15481 15520 PLECTIN 325.
FT REPEAT 15521 15560 PLECTIN 326.
FT REPEAT 15561 15600 PLECTIN 327.
FT REPEAT 15601 15640 PLECTIN 328.
FT REPEAT 15641 15680 PLECTIN 329.
FT REPEAT 15681 15720 PLECTIN 330.
FT REPEAT 15721 15760 PLECTIN 331.
FT REPEAT 15761 15800 PLECTIN 332.
FT REPEAT 15801 15840 PLECTIN 333.
FT REPEAT 15841 15880 PLECTIN 334.
FT REPEAT 15881 15920 PLECTIN 335.
FT REPEAT 15921 15960 PLECTIN 336.
FT REPEAT 15961 16000 PLECTIN 337.
FT REPEAT 16001 16040 PLECTIN 338.
FT REPEAT 16041 16080 PLECTIN 339.
FT REPEAT 16081 16120 PLECTIN 340.
FT REPEAT 16121 16160 PLECTIN 341.
FT REPEAT 16161 16200 PLECTIN 342.
FT REPEAT 16201 16240 PLECTIN 343.
FT REPEAT 16241 16280 PLECTIN 344.
FT REPEAT 16281 16320 PLECTIN 345.
FT REPEAT 16321 16360 PLECTIN 346.
FT REPEAT 16361 16400 PLECTIN 347.
FT REPEAT 16401 16440 PLECTIN 348.
FT REPEAT 16441 16480 PLECTIN 349.
FT REPEAT 16481 16520 PLECTIN 350.
FT REPEAT 16521 16560 PLECTIN 351.
FT REPEAT 16561 16600 PLECTIN 352.
FT REPEAT 16601 16640 PLECTIN 353.
FT REPEAT 16641 16680 PLECTIN 354.
FT REPEAT 16681 16720 PLECTIN 355.
FT REPEAT 16721 16760 PLECTIN 356.
FT REPEAT 16761 16800 PLECTIN 357.
FT REPEAT 16801 16840 PLECTIN 358.
FT REPEAT 16841 16880 PLECTIN 359.
FT REPEAT 16881 16920 PLECTIN 360.
FT REPEAT 16921 16960 PLECTIN 361.
FT REPEAT 16961 17000 PLECTIN 362.
FT REPEAT 17001 17040 PLECTIN 363.
FT REPEAT 17041 17080 PLECTIN 364.
FT REPEAT 17081 17120 PLECTIN 365.
FT REPEAT 17121 17160 PLECTIN 366.
FT REPEAT 17161 17200 PLECTIN 367.
FT REPEAT 17201 17240 PLECTIN 368.
FT REPEAT 17241 17280 PLECTIN 369.
FT REPEAT 17281 17320 PLECTIN 370.
FT REPEAT 17321 17360 PLECTIN 371.
FT REPEAT 17361 17400 PLECTIN 372.
FT REPEAT 17401 17440 PLECTIN 373.
FT REPEAT 17441 17480 PLECTIN 374.
FT REPEAT 17481 17520 PLECTIN 375.
FT REPEAT 17521 17560 PLECTIN 376.
FT REPEAT 17561 17600 PLECTIN 377.
FT REPEAT 17601 17640 PLECTIN 378.
FT REPEAT 17641 17680 PLECTIN 379.
FT REPEAT 17681 17720 PLECTIN 380.
FT REPEAT 17721 17760 PLECTIN 381.
FT REPEAT 17761 17800 PLECTIN 382.
FT REPEAT 17801 17840 PLECTIN 383.
FT REPEAT 17841 17880 PLECTIN 384.
FT REPEAT 17881 17920 PLECTIN 385.
FT REPEAT 17921 17960 PLECTIN 386.
FT REPEAT 17961 18000 PLECTIN 387.
FT REPEAT 18001 18040 PLECTIN 388.
FT REPEAT 18041 18080 PLECTIN 389.
FT REPEAT 18081 18120 PLECTIN 390.
FT REPEAT 18121 18160 PLECTIN 391.
FT REPEAT 18161 18200 PLECTIN 392.
FT REPEAT 18201 18240 PLECTIN 393.
FT REPEAT 18241 18280 PLECTIN 394.
FT REPEAT 18281 18320 PLECTIN 395.
FT REPEAT 18321 18360 PLECTIN 396.
FT REPEAT 18361 18400 PLECTIN 397.
FT REPEAT 18401 18440 PLECTIN 398.
FT REPEAT 18441 18480 PLECTIN 399.
FT REPEAT 18481 18520 PLECTIN 400.
FT REPEAT 18521 18560 PLECTIN 401.
FT REPEAT 18561 18600 PLECTIN 402.
FT REPEAT 18601 18640 PLECTIN 403.
FT REPEAT 18641 18680 PLECTIN 404.
FT REPEAT 18681 18720 PLECTIN 405.
FT REPEAT 18721 18760 PLECTIN 406.
FT REPEAT 18761 18800 PLECTIN 407.
FT REPEAT 18801 18840 PLECTIN 408.
FT REPEAT 18841 18880 PLECTIN 409.
FT REPEAT 18881 18920 PLECTIN 410.
FT REPEAT 18921 18960 PLECTIN 411.
FT REPEAT 18961 19000 PLECTIN 412.
FT REPEAT 19001 19040 PLECTIN 413.
FT REPEAT 19041 19080 PLECTIN 414.
FT REPEAT 19081 19120 PLECTIN 415.
FT REPEAT 19121 19160 PLECTIN 416.
FT REPEAT 19161 19200 PLECTIN 417.
FT REPEAT 19201 19240 PLECTIN 418.
FT REPEAT 19241 19280 PLECTIN 419.
FT REPEAT 19281 19320 PLECTIN 420.
FT REPEAT 19321 19360 PLECTIN 421.
FT REPEAT 19361 19400 PLECTIN 422.
FT REPEAT 19401 19440 PLECTIN 423.
FT REPEAT 19441 19480 PLECTIN 424.
FT REPEAT 19481 19520 PLECTIN 425.
FT REPEAT 19521 19560 PLECTIN 426.
FT REPEAT 19561 19600 PLECTIN 427.
FT REPEAT 19601 19640 PLECTIN 428.
FT REPEAT 19641 19680 PLECTIN 429.
FT REPEAT 19681 19720 PLECTIN 430.
FT REPEAT 19721 19760 PLECTIN 431.
FT REPEAT 19761 19800 PLECTIN 432.
FT REPEAT 19801 19840 PLECTIN 433.
FT REPEAT 19841 19880 PLECTIN 434.
FT REPEAT 19881 19920 PLECTIN 435.
FT REPEAT 19921 19960 PLECTIN 436.
FT REPEAT 19961 20000 PLECTIN 437.
FT REPEAT 20001 20040 PLECTIN 438.
FT REPEAT 20041 20080 PLECTIN 439.
FT REPEAT 20081 20120 PLECTIN 440.
FT REPEAT 20121 20160 PLECTIN 441.
FT REPEAT 20161 20200 PLECTIN 442.
FT REPEAT 20201 20240 PLECTIN 443.
FT REPEAT 20241 20280 PLECTIN 444.
FT REPEAT 20281 20320 PLECTIN 445.
FT REPEAT 20321 20360 PLECTIN 446.
FT REPEAT 20361 20400 PLECTIN 447.
FT REPEAT 20401 20440 PLECTIN 448.
FT REPEAT 20441 20480 PLECTIN 449.
FT REPEAT 20481 20520 PLECTIN 450.
FT REPEAT 20521 20560 PLECTIN 451.
FT REPEAT 20561 20600 PLECTIN 452.
FT REPEAT 20601 20640 PLECTIN 453.
FT REPEAT 20641 20680 PLECTIN 454.
FT REPEAT 20681 20720 PLECTIN 455.
FT REPEAT 20721 20760 PLECTIN 456.
FT REPEAT 20761 20800 PLECTIN 457.
FT REPEAT 20801 20840 PLECTIN 458.
FT REPEAT 20841 20880 PLECTIN 459.
FT REPEAT 20881 20920 PLECTIN 460.
FT REPEAT 20921 20960 PLECTIN 461.
FT REPEAT 20961 21000 PLECTIN 462.
FT REPEAT 21001 21040 PLECTIN 463.
FT REPEAT 21041 21080 PLECTIN 464.
FT REPEAT 21081 21120 PLECTIN 465.
FT REPEAT 21121 21160 PLECTIN 466.
FT REPEAT 21161 21200 PLECTIN 467.
FT REPEAT 21201 21240 PLECTIN 468.
FT REPEAT 21241 21280 PLECTIN 469.
FT REPEAT 21281 21320 PLECTIN 470.
FT REPEAT 21321 21360 PLECTIN 471.
FT REPEAT 21361 21400 PLECTIN 472.
FT REPEAT 21401 21440 PLECTIN 473.
FT REPEAT 21441 21480 PLECTIN 474.
FT REPEAT 21481 21520 PLECTIN 475.
FT REPEAT 21521 21560 PLECTIN 476.
FT REPEAT 21561 21600 PLECTIN 477.
FT REPEAT 21601 21640 PLECTIN 478.
FT REPEAT 21641 21680 PLECTIN 479.
FT REPEAT 21681 21720 PLECTIN 480.
FT REPEAT 21721 21760 PLECTIN 481.
FT REPEAT 21761 21800 PLECTIN 482.
FT REPEAT 21801 21840 PLECTIN 483.
FT REPEAT 21841 21880 PLECTIN 484.
FT REPEAT 21881 21920 PLECTIN 485.
FT REPEAT 21921 21960 PLECTIN 486.
FT REPEAT 21961 22000 PLECTIN 487.
FT REPEAT 22001 22040 PLECTIN 488.
FT REPEAT 22041 22080 PLECTIN 489.
FT REPEAT 22081 22120 PLECTIN 490.
FT REPEAT 22121 22160 PLECTIN 491.
FT REPEAT 22161 22200 PLECTIN 492.
FT REPEAT 22201 22240 PLECTIN 493.
FT REPEAT 22241 22280 PLECTIN 494.
FT REPEAT 22281 22320 PLECTIN 495.
FT REPEAT 22321 22360 PLECTIN 496.
FT REPEAT 22361 22400 PLECTIN 497.
FT REPEAT 22401 22440 PLECTIN 498.
FT REPEAT 22441 22480 PLECTIN 499.
FT REPEAT 22481 22520 PLECTIN 500.
FT REPEAT 22521 22560 PLECTIN 501.
FT REPEAT 22561 22600 PLECTIN 502.
FT REPEAT 22601 22640 PLECTIN 503.
FT REPEAT 22641 22680 PLECTIN 504.
FT REPEAT 22681 22720 PLECTIN 505.
FT REPEAT 22721 22760 PLECTIN 506.
FT REPEAT 22761 22800 PLECTIN 507.
FT REPEAT 22801 22840 PLECTIN 508.
FT REPEAT 22841 22880 PLECTIN 509.
FT REPEAT 22881 22920 PLECTIN 510.
FT REPEAT 22921 22960 PLECTIN 511.
FT REPEAT 22961 23000 PLECTIN 512.
FT REPEAT 23001 23040 PLECTIN 513.
FT REPEAT 23041 23080 PLECTIN 514.
FT REPEAT 23081 23120 PLECTIN 515.
FT REPEAT 23121 23160 PLECTIN 516.
FT REPEAT 23161 23200 PLECTIN 517.
FT REPEAT 23201 23240 PLECTIN 518.
FT REPEAT 23241 23280 PLECTIN 519.
FT REPEAT 23281 23320 PLECTIN 520.
FT REPEAT 23321 23360 PLECTIN 521.
FT REPEAT 23361 23400 PLECTIN 522.
FT REPEAT 23401 23440 PLECTIN 523.
FT REPEAT 23441 23480 PLECTIN 524.
FT REPEAT 23481 23520 PLECTIN 525.
FT REPEAT 23521 23560 PLECTIN 526.
FT REPEAT 23561 23600 PLECTIN 527.
FT REPEAT 23601 23640 PLECTIN 528.
FT REPEAT 23641 23680 PLECTIN 529.
FT REPEAT 23681 23720 PLECTIN 530.
FT REPEAT 23721 23760 PLECTIN 531.
FT REPEAT 23761 23800 PLECTIN 532.
FT REPEAT 23801 23840 PLECTIN 533.
FT REPEAT 23841 23880 PLECT
```

mapping of a single phosphorylation site.";

RL J. Biol. Chem. 271:8203-8208(1996).

CC -!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND  
 CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR  
 CC HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND  
 CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT  
 CC ALSO IN THE REGULATION OF THEIR DYNAMICS.

CC -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.

CC -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH  
 CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-  
 CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.

CC -!- PTM: PHOSPHORYLATION BY CDC2; REGULATES DISSOCIATION FROM  
 CC INTERMEDIATE FILAMENTS DURING MITOSIS.

CC -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.

CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.

CC -!- SIMILARITY: CONTAINS 32 PLECTIN REPEATS.

CC -!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.

CC -!- SIMILARITY: BELONGS TO THE PLAGIN OR CYTOLINKER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF260753; AAF70372.1;

DR HSSP; O01082; 1BKR.

DR InterPro: IPR001589; Actbind\_actinin.

DR InterPro: IPR001715; Calponin-like.

DR InterPro: IPR001101; Plectin\_repeat.

DR InterPro: IPR002017; Spectrin.

DR Pfam: PF00307; CH; 2.

DR Pfam: PF00681; Plectin; 20.

DR SMART; SM00033; CH; 2.

DR SMART; SM00250; PLEC; 32.

DR SMART; SM00150; SPEC; 4.

DR PROSITE; PS00019; ACTININ\_1; PARTIAL.

DR PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.

DR PROSITE; PS00021; CH; 2.

KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;

KW Phosphorylation.

FT NON\_TER 1

FT DOMAIN <1 1259

FT DOMAIN 1260 2544

FT DOMAIN 2545 4473

FT DOMAIN <1 192

FT DOMAIN <1 174

FT DOMAIN 87 189

FT REPEAT 449 508

FT REPEAT 529 613

FT REPEAT 626 719

FT REPEAT 1104 1204

FT REPEAT 1258 2548

FT REPEAT 2615 2652

FT REPEAT 2653 2690

FT REPEAT 2691 2728

FT REPEAT 2729 2766

FT REPEAT 2770 2804

FT REPEAT 2905 2942

FT REPEAT 2943 2980

FT REPEAT 2981 3018

FT REPEAT 3019 3056

FT REPEAT 3057 3094

FT REPEAT 3274 3311

FT REPEAT 3312 3349

FT REPEAT 3350 3387

FT REPEAT 3388 3425

FT REPEAT 3429 3463

FT REPEAT 3609 3646

FT REPEAT 3647 3684

FT REPEAT 3685 3722

FT REPEAT 3723 3760

FT REPEAT 3764 3797

FT REPEAT 3800 3834

FT REPEAT 3852 3889

FT REPEAT 3890 3927

FT REPEAT 3928 3965

FT REPEAT 3966 4003

FT REPEAT 4007 4041

FT REPEAT 4043 4094

FT REPEAT 4197 4234

FT REPEAT 4235 4272

FT REPEAT 4273 4310

FT REPEAT 4311 4348

FT REPEAT 4349 4386

FT REPEAT 4399 4089

FT DOMAIN 4414 4429

FT MOD\_RES 4328 4328

SQ SEQUENCE 4473 AA; 509015 MW; E14615D361E3484 CRC64;

Query Match 9.0%; Score 328; DB 1; Length 4473;  
 Best Local Similarity 23.8%; Pred. No. 6.5e-07;  
 Matches 187; Conservative 116; Mismatches 274; Indels 210; Gaps 32;

QY 65 WKLYSNLKWLRSLVEYSQDVLAH-PVSEELPDVS-----LIGFSDPAELGKLLQLVL 119  
 DB 1103 WREVN-----OLLERQWAVLAQIDVQRELEGRQLRYRESADP--LSSWQ---- 1150

QY 120 GCAISCEKQDHIQIMTLEESVQHVMEAIQELMTKDTPSLSPTYGNFDSQSRYYF 179  
 DB 1151 -----DAKRQEQIQAV-----PIPNQAAAR-----EQLRQKA 1179

QY 180 LSEAEPEGDELOQC-----LDLERQLMLSEE-----KOSLA 212  
 DB 1180 LLEETERHGEKVEECOKFAKQVINAIKDYELQVLYKAQLEPVASPAKKPKVQSGSEVI 1239

QY 213 QENAGLRERMRPEGEGTGLTAKKLLLLQSLEQLQEEFNFLSCREDERLCAP----- 268  
 DB 1240 QEYVQLTRYSE-----LTLTISQYIKFISETLRRMEEERLAEQORAEERELAGEAA 1294

QY 269 -----LEREVAELQHRNQ-----ALTSLAQAQAKLKDMDLRSQ 303  
 DB 1295 LEKQROLAEHAQAQAQAELEAQELQRRWQEEVAREEAAVNAQQQKRSIQEELHRLQS 1354

QY 304 SE-----RAGQLEATLTSRRRLGELRELRRQVRQLERNAHAETROLEDELRLAGSL 358  
 DB 1355 SEAEIQAKAQVEAERSRMRIEERIVVRVRLQLETTERQGG-----AEGELQALRAR 1407

QY 359 RAQLEAQRQVQE-----LOGQROEAMKAKFWLFEGRNLEEKYESVTKEKERLLAERDSL 414  
 DB 1408 AEEAEAKQKQAEERLRRQVQDESQKKQAEALALRVKQAQAEAKQKRALQALEEL 1467

QY 415 R-----EANEELRCAQLQPRGLTQADPSLDSTPTSPVONLAAEILPAE-----LRE--- 459  
 DB 1468 RLQAEAEERLRQAEARARQVQVALETAQPSAEVQLSKRASFQAEKTAQLETLQSEHV 1527

QY 460 TLLRLQLENKRLCRQEAADRERQELQHLH-----DANRA-RHGLETOHRLNQOQSELR 514  
 DB 1528 TVTQLREKAEERRAQQAQAEARAEAREERELERWQLKANEALRLRLQAEVAQQRSLAQAD 1587

QY 515 A--QVEDLQKALQEQGKTEDAISLLKRL--BEHLQKLHE-----ADLEL--- 557  
 DB 1588 AEKQKEAEAREARRGKAQEAQR--ORELAQEELKQRLAEGTAQQAQRLAAEQELIRL 1644

QY 558 -----QRKREYIEE-----LEPPTDSTARRTEELQHNKQKDDADLRAM-----EER 599  
 DB 1645 RAETQEQEQORQLLEELARLQREANTATHKR-OELEPAELAKVRAENEVLASKARAE 1703

QY 600 YRYVDKARVMQTMPEKQKPAAGAPPELHSL-----RTQLRERDVIRIHLMD----- 648  
 DB 1704 SRSTSEKSKORLEAEDRFRELAEEAARLALAEAKRQRLAEADAARQAEAEVILTE 1763

QY 649 -----FEKRSRQEQEKKLLISAWNMGM-----ALQQRAGEERAPAH---AQSFLAQ 693

Db 1764 KLAISEATRLKTEATKKEAENRRLRLAEAPQRRRLLEQAALHADIERLAQ 1823  
 QY 694 ORLATNS 700  
 Db 1824 LRKASES 1830  
 RESULT 5  
 MYHA\_BOVIN STANDARD; PRT; 1976 AA.  
 ID MYHA\_BOVIN STANDARD; PRT; 1976 AA.  
 AC Q27991; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE MYOSIN heavy chain, nonmuscle type B (Cellular myosin heavy chain,  
 type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).  
 GN MYH10.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ohara M., Ishiguro N., Shinagawa M.;  
 RT "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 204-302 FROM N.A.  
 RC TISSUE=Brain cortex;  
 RX MEDLINE=95301542; PubMed=7782316;  
 RA Itoh K., Adelstein R.S.;  
 RT "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle  
 myosin heavy chain II-B".  
 RL J. Biol. Chem. 270:14533-14540(1995).  
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,  
 CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND  
 CAPPING (BY SIMILARITY).  
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY  
 CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2  
 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC ENBL; AB022023; BAA36494.1;  
 CC EMBL; U15716; AAA87715.1;  
 CC HSSP; P10587; 1BR2.  
 DR InterPro; IPR000048; IQ\_region.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR InterPro; IPR002017; Spectrin.  
 DR InterPro; IPR001609; myosin\_head.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR ProDom; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR PROSITE; PS50096; IQ; 1.

KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;  
 KW Coiled coil; Alkylation; Multigene family.  
 FT DOMAIN 1 785 MYOSIN HEAD-LIKE.  
 FT DOMAIN 786 815 IQ.  
 FT DOMAIN 845 1976 COILED COIL (POTENTIAL).  
 FT NP\_BIND 178 185 ATP (POTENTIAL).  
 FT MOD\_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).  
 FT MOD\_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).  
 SQ SEQUENCE 1976 AA; 229097 MW; 6144354451C0F790 CRC64;  
 Query Match 8.8%; Score 321; DB 1; Length 1976;  
 Best Local Similarity 24.8%; Pred. No. 5.6e-07;  
 Matches 174; Conservative 120; Mismatches 286; Indels 122; Gaps 28;  
 QY 101 LIGFSD--PABLGKLLQLVLGCAISCEKKQDHIQIRIM-----TLESVQHVVMETIQEL 153  
 Db 1060 LDGETTLDQOIAELQAIQIDELKIQVAKKEEELGALARGDDETLHKNNALKVVRQLQAQ 1119  
 QY 154 MTKPTPSLSPTYGNDSDSRRYFELSEAE--GDELOQRCLDLERQLMLLSSEKQSLA 212  
 Db 1120 IAELOQDFESEKASRNKAOKR--DISEELEAKTELDLTDTAAQQELURKQGEVA 1177  
 QY 213 QENAGLRMRGREG--EGTPGLTAKKLLLLQSQLEQLQEENFRLESGR-----EDERLR 265  
 Db 1178 ELKKALEETKSHQAQIDMRQRHATALEELSELEQAQKRFKANLEKNKQGLETDNKELA 1237  
 QY 266 C--AELEREVAELQHRNQALTSLAQEAQALKDMDERL--QSSERAGOLEATLTSCRRRLG 322  
 Db 1238 CEVKVQOVKAESHKRKKLDAQVOELHAKVSEGDRRLVELAEKANKLQNELDNLVSTLE 1297  
 QY 323 E-----LRELRRQ-----VROLEERNAGHAERTROLED 350  
 Db 1298 EAEKKGIFKADANGLESQLODTQELQETTRQKLNLSRRIRLEERSLQEQOEE--EE 1356  
 QY 351 ELRRA-----GSLRAQLEAQRQVQELQG--QROEAMKAQKWLFE-----CRNLEEK--- 396  
 Db 1357 EARSLEKQLQALQALQALTDTKKVDLDLGTIENLEAKK--KLLKQVEVLSORLEEKALA 1414  
 QY 397 YESVTKERKELLARDSL-----REANEELRCAQLQPRGLTQADPSLDPT--STPDVNL 448  
 Db 1415 YDKLETKTRLOQELDLVLDHQROIVSNLEKKQKFKDQLLAEKKNISARYAEERDRA 1474  
 QY 449 AAEILPAELRETLRLQENKRLCRQEAADRQ-----BELQHLLEDANRARGHLETOH 503  
 Db 1475 EAERARETKALSARALEALEAREAEQRNQLRADMDLMSKDDVGNKVHVELEKSK 1534  
 QY 504 RLNQOQLSELRAQVEDLQALQEQGGTDAISILLKRLKEEHQKLH--EADLELQKRE 562  
 Db 1535 RALEQQQVEEMRTQLELEDELQ-----ATEDA-----KLRLVNNQAMKAQFERDLQTRDE 1585  
 QY 563 YIELEPPTSSSTARRIEELQHNLOKQKDALRAMEERYRYV-----DKARMVMQTMPEK 617  
 Db 1586 QNEKK-----RLLIKQVRELEAELE-----DERKQALAVASKKMEIDLKLEAQ 1632  
 QY 618 QRPAAQAPPELHSLRDLRDRVIRHLEMDFEKSSORE-----QEEKLLISAWNYM 670  
 Db 1633 IEANKARDEVK---OLRKLQAMQKDYQRELEEARSDREIFAQSESEKSKLKLSLEAEI 1689  
 QY 671 GMAQQORAGEERAPAHQAQSLAQORLATNSRRGPLGLASLN 712  
 Db 1690 LQLEELASSERARRHAQ--ERDELADELTANSASGKSALLD 1729  
 RESULT 6  
 MYH9\_RAT  
 ID MYH9\_RAT STANDARD; PRT; 1961 AA.  
 AC Q62812;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,  
 type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).  
 GN MYH9.

OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus  
 OX NCBI\_TaxID=10116;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,  
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND  
 CC CAPPING.  
 CC  
 CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY  
 CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2  
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC  
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC  
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U31463; AAA74950.1;  
 DR HSP: P10587; IBL2;  
 DR InterPro: IPR000048; IQ\_region.  
 DR InterPro: IPR004009; Myosin\_N.  
 DR InterPro: IPR002928; Myosin\_tail.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR Pfam: PF00812; IQ; 1.  
 DR Pfam: PF01576; Myosin\_tail; 1.  
 DR Pfam: PF02736; Myosin\_N; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head; 1.  
 DR SMART: SM00015; IQ; 1.  
 DR SMART: SM00242; MYSC; 1.  
 DR PROSITE: PS50096; IQ; 1.  
 DR KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;  
 KW Coiled coil; Alkylation; Multigene family.  
 FT DOMAIN 1 778  
 FT MYOSIN HEAD-LIKE.  
 FT  
 FT DOMAIN 779 808  
 FT IQ.  
 FT  
 FT DOMAIN 841 1927  
 FT COILED COIL (POTENTIAL).  
 FT NP\_BIND 174 181  
 FT ATP (POTENTIAL).  
 FT DOMAIN 654 676  
 FT ACTIN-BINDING.  
 FT MOD\_RES 694 694  
 FT ALKYLATION (SH-1) (POTENTIAL).  
 FT MOD\_RES 704 704  
 FT ALKYLATION (SH-2) (POTENTIAL).  
 FT SEQUENCE 1961 AA; 226336 MW; 989876D9681FB19E CRC64;  
 Query Match 8.8%; Score 320.5; DB 1; Length 1961;  
 Best Local Similarity 23.7%; Pred. No. 5.8e-07;  
 Matches 179; Conservative 133; Mismatches 269; Indels 175; Gaps 33;  
 QY 54 LOGISEDGP---PNWKLKVNKLWLSVYSQDVLAHPVSEHLPDVSIL--IGEP-SD 107  
 DB 1053 LEGDSTLSQIAELAQIAELAKMQLAKKEELQAAARVEEAQAQNMALKKIRELETO 1112  
 QY 108 PAELGKLLQVLGCAISCEKQ-----DHQIRMT-LEESVQHVVMEAIQELMTKDTPD-S 161  
 DB 1113 ISELQEDLESERACRNKAEKQKRDGLGELEALKTELEDTLDTA--AQQLSRKREQVS 1170  
 QY 162 LSPETQNFDSQSRFYFLSE-----AEGDELOQRCCLDLERQMLLSEKQSLAENA 216  
 DB 1171 ILKKT---LEDEANTHEAQIQEMKQKQSOAVEELAEOLEQTKRVKATLEKAKOTLENERG 1227  
 QY 217 GL-RERMGRPEGTPGTAKLLQLLQSQLOEQENFRLESGREDEFLRCAELEREVAE 275  
 DB 1228 ELANEVKALQKGDSEHKRKK---VEAQLOELQVK-----FSEGERVVR-TELADRVSK 1277

QY 276 LOHRNOALTSIAQEAQALKDEMDELQSSERAGOLEATLTSCRRRLGELRELRRQ----- 330  
 DB 1278 LQVELDSVTGL-----LNQSDSKSKLTDFSALESQLODTQELLQNRQK 1324  
 QY 331 -----VQLEERNAGHAERTROLEDELRR-----AGSLRAQ 361  
 DB 1325 LSLSTKLKQMEDEKNSFREQLLEEEAEAKRLKQIATLHAQVTDMMKKMDEGVCCLETA 1384  
 QY 362 LEAQRRQVQELQ--QRQEEAMKAELWLFECRNLEEKYESVTKEKRLLAERDSLRANE 419  
 DB 1385 EAKRRRLQKDLGSLQRLEKVA-----YDKLEKTKRLQOELDDLVLVDLD 1431  
 QY 420 ELRCAQLPRLGTQADPSLDPTSTPVDNLAAE--ILPAELRETLRLQLENKR----- 470  
 DB 1432 HQR-----QSVNLEKKQKFDQLAAEKTISAKYAEERDRAEAAREKETKALS 1481  
 QY 471 LCR-----QEAADRER-----QEBLQRHLEDANRARGHLETOHRLNQOQLSELRAQ 516  
 DB 1482 LARALEEAMEQKAELEKLNKQFRTMEDLMSSKDDVGKSVHELEKSNRALEQQOQVEEMKTQ 1541  
 QY 517 VEDQKALQEQGGKTEDAISILLKKLEHLOKLH-EADLELQKREYIEELEPTDSST 575  
 DB 1542 LEELEDELQ-----APEDA-----KLRLVNQAMKAQAFERDLOGRDEQSEEEKK-----KOL 1588  
 QY 576 ARRIEELQHNQKQKADLRAEMEERYRVYVDKARVMQTMPEKQKQPA--AGAPPELHSLRT 633  
 DB 1589 VQVREMEAELE-----DERQKRSIAMAARKKLEMDLKDLEAHIDTANKNREEAIK 1639  
 QY 634 QLRERDVRIHLEMDFEKRSQRE-----QEEKLLISAWYNMGMALQOQAGEERAPAH 686  
 DB 1640 QLRKLAQMKDCMRDVEDTFRASREELAQAKENKRLKLSMEAEIQLQBELAAERAKRQ 1699  
 QY 687 AQSFLAQQR-----LATNSRRGPIG-----RLASL 711  
 DB 1700 AQ-----QERDELADEIANSSGKALALEKRRLEAL 1731  
 RESULT 7  
 ID TRHY\_SHEEP STANDARD; PRT: 1549 AA.  
 AC P22793;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trichohyalin.  
 GN THH.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93260018; PubMed=7684041;  
 RA Fietz M.J., McLaughlan C.J., Campbell M.T., Rogers G.E.;  
 RT "Analysis of the sheep trichohyalin gene: potential structural and  
 RT calcium-binding roles of trichohyalin in the hair follicle."  
 RL J. Cell Biol. 121:855-865(1993).  
 RN [2]  
 RP SEQUENCE OF 1016-1549 FROM N.A.  
 RC STRAIN=Merino-Dorset horn X Border Leicester; TISSUE=wool follicles;  
 RX MEDLINE=90130632; PubMed=2298812;  
 RA Fietz M.J., Presland R.B., Rogers G.E.;  
 RT "The cDNA-deduced amino acid sequence for trichohyalin, a  
 RT differentiation marker in the hair follicle, contains a 23 amino acid  
 RT repeat."  
 RL J. Cell Biol. 110:427-436(1990).  
 CC -!- FUNCTION: INTERMEDIATE KERATIN-ASSOCIATED PROTEIN THAT ASSOCIATES  
 CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
 CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER



WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL DIFFERENTIATION.

-1- SUBUNIT: HOMODIMER (PROBABLE).

-1- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS MAY BE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.

-1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN THE EPITHELIA OF THE TONGUE, HOOF AND RUMEN.

-1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS. DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG DIFFERENT SPECIES.

-1- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.

-1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100 FAMILY.

-1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Z18361; CA79165.1; -  
EMBL; X51695; CA33592.1; -  
PIR; A34209; A34209.  
PIR; S32633; S32633.  
PIR; A40691; A40691.  
HSSP; P02633; 1IG5.  
InterPro; IPR001751; CABP\_S100.  
InterPro; IPR002048; EF-HAND.  
Pfam; PF00036; ehand; 1.  
Pfam; PF01023; S\_100; 1.  
ProDom; PD003407; CABP\_S100; 1.  
ProSITE; PS00018; EF\_HAND; 1.  
ProSITE; PS00303; S100\_CABP; FALSE\_NEG.  
Keratinization; Repeat; Calcium-binding.  
DOMAIN 1 91  
CA\_BIND 22 33  
EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
CA\_BIND 62 73  
EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).  
DOMAIN 413 832  
EF-HAND 14 X 28 AA APPROXIMATE TANDEM REPEATS.  
REPEAT 413 448 1-1.  
REPEAT 449 476 1-2.  
REPEAT 477 504 1-3.  
REPEAT 505 532 1-4.  
REPEAT 533 560 1-5.  
REPEAT 561 588 1-6.  
REPEAT 589 616 1-7.  
REPEAT 617 644 1-8.  
REPEAT 645 678 1-9.  
REPEAT 679 706 1-10.  
REPEAT 707 742 1-11.  
REPEAT 743 771 1-12.  
REPEAT 772 796 1-13.  
REPEAT 797 832 1-14.  
DOMAIN 938 1507 23 X 23 AA APPROXIMATE TANDEM REPEATS.  
REPEAT 938 961 2-1.  
REPEAT 962 985 2-2.  
REPEAT 986 1021 2-3.  
REPEAT 1022 1044 2-4.  
REPEAT 1045 1067 2-5.  
REPEAT 1068 1090 2-6.  
REPEAT 1091 1121 2-7.

FT REPEAT 1122 1144 2-8.  
FT REPEAT 1145 1167 2-9.  
FT REPEAT 1168 1197 2-10.  
FT REPEAT 1198 1227 2-11.  
FT REPEAT 1228 1250 2-12.  
FT REPEAT 1251 1273 2-13.  
FT REPEAT 1274 1296 2-14.  
FT REPEAT 1297 1319 2-15.  
FT REPEAT 1320 1342 2-16.  
FT REPEAT 1343 1368 2-17.  
FT REPEAT 1369 1391 2-18.  
FT REPEAT 1392 1416 2-19.  
FT REPEAT 1417 1439 2-20.  
FT REPEAT 1440 1461 2-21.  
FT REPEAT 1462 1484 2-22.  
FT REPEAT 1485 1507 2-23.  
FT VARIANT 1145 1197 MISSING (IN SHORT FORM).  
FT VARIANT 1251 1273 MISSING (IN SHORT FORM).  
FT CONFLICT 1399 1399 E -> G (IN REF. 2).  
SQ SEQUENCE 1549 AA; 201173 MW; E72PB9FF1326E54E CRC64;

Query Match 8.8%; Score 319; DB 1; Length 1549;  
Best Local Similarity 23.5%; Pred. No. 5.3e-07;  
Matches 172; Conservative 104; Mismatches 241; Indels 216; Gaps 28;

QY 126 EKKQDHQIRIMTLEESVQHVMEAIQELMTKPTDPSLSPTYGNFD-----SQSR 175  
DB 319 QEOQSRQOELEERQOEIOISEVQSLQEDQGRQLKQEQ--RYDQNRWQLEESQRR 376  
QY 176 RYVFLS-----EEAEGDEL-----OQCLDLERQLMLLSEKQSLA 212  
DB 377 RYLYAKPAQRQVREERQRLKEEKQREKQREKQREKQREKQREKQREKQREKQRE 436  
QY 213 QENAGLRERMGPGEPTGLTAKKLLLSQLEQLEENFRLESG-----REDER 263  
DB 437 REE--REKRRQERE--KQYLEKVELNEEQLOREERKQREKQREKQYLEKVELNEEQ 491  
QY 264 LRCAELERVAELQHRNQALTSIAQEAQALKDEMDELROSSER-----AGOLEAT 313  
DB 492 LQOERERKQREKQREKQYLEKVELNEEQLOREERKQREKQREKQREKQREKQREKQ 551  
QY 314 LITSCRR-----LGLERLRQVRQ--LEERNAGHAERTR-OLEDELRRAGSLR 359  
DB 552 EREKRRQREKQYLEKVELNEEQLOREERKQREKQREKQREKQYLEKVELNEEQLOREER 611  
QY 360 AOLEAQRQVQELQQRQEAAMKAELFECNLEEKYESVTKE-----KERLIAERDSL 415  
DB 612 RQERERQYLEKVELNEEQVQQRQER--EKRRQERQYLEKELQQRQERQEEQLLR 668  
QY 416 EANEELRCAQLQPRGLTQADPSLDPTSTPDVNLAAHILPAELRETLRLQLNKLRCROE 475  
DB 669 EREKRR--QERERQYLE-----KVELQEE--EQLOREERKRRQE 705  
QY 476 AADRERQ-----EELQRHLEDANRARGHLETOHRLNQOQLSELRAQVEDLQ----- 521  
DB 706 ---REKQYLEKVELNEEQLOREERKQREKQREKQREKQREKQREKQREKQREKQ 762  
QY 522 KALQOQGGKTDAISILLKRLKEHLQ-----KLHEADLELQR 559  
DB 763 QLLREDREKQYLEKVELQRE--EQLOREKQREKQREKQREKQREKQREKQREKQ 821  
QY 560 KREYIELEPPTDSSTARREE-----LQHLNKKKADLRAMEERYRRYVDKAR----- 608  
DB 822 ----EECEKRRQRELEERQLEELQRLDRKKQFRDDDDQHONVRSVYKURENKES 876  
QY 609 -----MYQTMPE-----KORPAAGAPPELHSLRQLRE----- 637  
DB 877 ROLDSDSWRESOFQDLRLQDEQREKREKREKREKREKREKREKREKREKREKREKRE 936  
QY 638 -----RDVEIRHLEMDFEKRSQRQREKREKREKREKREKREKREKREKREK 682  
DB 937 RDRKFRREEQLLKQREKIRYLEED-----RKFREEQOL-----RRLEREQQLRQER 985



QY 583 APAHQSFQAQR 695  
DB 986 DRKFRELSROER 998

RESULT 8

TRH1\_HUMAN STANDARD; PRT: 1898 AA.

AC Q07283;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Trichohyalin.

GN THH OR THRY OR THL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID:9606;

RA (1)

RP SEQUENCE FROM N.A.

RX MEDLINE-93280194; PubMed-7685034;

RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,

RA Steinert P.M.;

RT "The structure of human trichohyalin. Potential multiple roles as a

RT functional EF-hand-like calcium-binding protein, a cornified cell

RT envelope precursor, and an intermediate filament-associated (cross-

RT linking) protein.";

RL J. Biol. Chem. 268:12164-12176(1993).

RN (2)

RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.

RX MEDLINE-93315897; PubMed-7686953;

RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;

RT "Trichohyalin: a structural protein of hair, tongue, nail, and

RT epidermis.";

RL J. Invest. Dermatol. 101:655-715(1993).

CC (1) FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES

CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE

CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR

CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY

CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER

CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN

CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN

CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL

CC DIFFERENTIATION.

CC (1) SUBUNIT: MONOMER (PROBABLE).

CC TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS

CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN

CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).

CC DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF

CC THE EPIDERMIS.

CC (1) DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND

CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST

CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS

CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED

CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS

CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.

CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN

CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG

CC DIFFERENT SPECIES.

CC (1) PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE

CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.

CC (1) SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100

CC FAMILY.

CC (1) SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

EMBL; L09190; AAA65582.1;  
PIR; A45973; A45973.  
HSP; P02633; 4ICB.  
Genew; HGNC:11791; THH.  
MIM; 190370;  
InterPro; IPR001751; CaBP\_S100.  
InterPro; IPR002048; EF-hand.  
InterPro; IPR002017; Spectrin.  
Pfam; PF00036; efhand; 1.  
Pfam; PF01023; S\_100; 1.  
ProDom; PD003407; CaBP\_S100; 1.  
ProSITE; PS00018; EF\_HAND; 1.  
ProSITE; PS00303; S100\_CaBP; 1.  
Keratinization; Repeat; Calcium-binding.  
FT DOMAIN 1 31  
FT CA\_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
FT CA\_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).  
FT DOMAIN 314 390 6 X 13 AA TANDEM REPEATS OF  
R-R-E-Q-E-E-R-R-E-Q-Q-L.  
FT REPEAT 314 326 S-100 LIKE.  
FT REPEAT 327 339 1-1 (APPROXIMATE).  
FT REPEAT 340 351 1-2 (APPROXIMATE).  
FT REPEAT 352 364 1-3 (APPROXIMATE).  
FT REPEAT 365 377 1-4.  
FT REPEAT 378 390 1-5.  
FT DOMAIN 391 444 1-6.  
FT REPEAT 391 396 2-1.  
FT REPEAT 397 402 2-2.  
FT REPEAT 403 408 2-3.  
FT REPEAT 409 414 2-4.  
FT REPEAT 415 420 2-5.  
FT REPEAT 421 426 2-6.  
FT REPEAT 427 432 2-7.  
FT REPEAT 433 438 2-8.  
FT REPEAT 439 444 2-9.  
FT DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDEM REPEATS.  
FT DOMAIN 923 1162 8 X 30 AA TANDEM REPEATS.  
FT REPEAT 923 952 4-1.  
FT REPEAT 953 982 4-2.  
FT REPEAT 983 1012 4-3.  
FT REPEAT 1013 1042 4-4.  
FT REPEAT 1043 1072 4-5.  
FT REPEAT 1073 1102 4-6.  
FT REPEAT 1103 1132 4-7.  
FT REPEAT 1133 1162 4-8.  
FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.  
FT CONFLICT 1752 1752 F -> L (IN REF. 2).  
FT CONFLICT 1794 1801 QERDQRYR -> RSETGSTG (IN REF. 2).  
FT CONFLICT 1857 1857 Q -> K (IN REF. 2).  
FT CONFLICT 1880 1880 V -> G (IN REF. 2).  
SQ SEQUENCE 1898 AA; 247219 MW; A74B5947FB62E31D CRC64;

Query Match 8.7%; Score 315; DB 1; Length 1898;

Best Local Similarity 24.5%; Pred. No. 9.7e-07;

Matches 168; Conservative 110; Mismatches 269; Indels 140; Gaps 29;

QY 104 EFSDAELCK-LLQL-VLGCALCEKKDHIQRIINTLESVQ-----HVVMEAIQELMTK 156  
DB 203 EFPDESQRLRELLLELRKRGEEKQOORRQDRVFQEEEEKRWKRRVTLRKEEKLOE 262  
QY 157 DTPD-----SLSPITYGNFDSQSRRY--YFLSEAEAEDELQQRCLDLRLMLLSEKQSL 211  
DB 263 EFPQORELOEE-----EQLRLKLEQLRRERQEEQQOOR-LRREQLRRKQEEERRE 316  
QY 212 AGENAGLRMRGPEGEGTGLTAKKLLLLQSQLEQLQENFLESGRDERURCAELER 271  
DB 317 QOER--RQOERRE-----QOERREQLR-----RQOERREQLRR 353  
QY 272 EVELQHRNQLTSLAQE---AQAALKEDELQSSERAGLEATLTSRRRLGELR--- 325  
DB 354 EQEE-ERREQLRRQOEEERREQLRRQOEEERREQLRRQOEEERREQLRRQOEEERRE 412  
QY 326 ELRRVQRLERENAGHARTQLEDELRLRAGLSLAQLEAQRQVQELQGGQEEAMKAEK 385



CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

-1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES. CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

-1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGLIN ANOMALY (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.

-1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF FECHTNER SYNDROME (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS. WITH ADDITIONAL ALPORT-LIKE CLINICAL FEATURES OF SENSORINEURAL DEAFNESS, CATARACTS AND NEPHRITIS.

-1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.

-1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA17) WHICH IS CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND COCHLEOSACCULAR DEGENERATION.

-1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

-1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; 282215; CAB05105.1; -  
 DR EMBL; M81105; AAA59888.1; -  
 DR EMBL; M69180; AAA61765.1; -  
 DR EMBL; M31013; AAA36349.1; -  
 DR HSP; P10587; 1B2.  
 DR Genew; HGNC:7579; MYH9.  
 DR MIM; 160775; -  
 DR MIM; 153640; -  
 DR MIM; 155100; -  
 DR MIM; 603622; -  
 DR MIM; 605249; -  
 DR InterPro; IPR000048; IQ\_region.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR InterPro; IPR002017; Spectrin.  
 DR InterPro; IPR001609; myosin\_head.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR ProDom; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR PROSITE; PS00936; IQ; 1.  
 DR PROSITE; PS00936; IQ; 1.  
 KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;  
 KW Coiled coil; Alkylation; Multigene family; Disease mutation;  
 KW Deafness.  
 FT DOMAIN 1 778 MYOSIN HEAD-LIKE.  
 FT DOMAIN 779 808 IQ.  
 FT DOMAIN 837 1926 COILED COIL (POTENTIAL).  
 FT NP\_BIND 174 181 ATP (POTENTIAL).  
 FT DOMAIN 654 676 ACTIN-BINDING.  
 FT MOD\_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).  
 FT MOD\_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).  
 FT VARIANT 93 93 N -> K (IN MHA).  
 FT VARIANT 702 702 R -> C (IN FTNS).  
 FT VARIANT 705 705 R -> H (IN FTNS).  
 FT VARIANT 1155 1155 R -> H (IN DFNA17).  
 FT VARIANT 1155 1155 T -> I (IN MHA).

FT VARIANT 1165 1165 /FTID=VAR\_010794.  
 FT VARIANT 1424 1424 R -> C (IN SBS).  
 FT VARIANT 1841 1841 /FTID=VAR\_010795.  
 FT VARIANT 1841 1841 D -> H (IN FTNS).  
 FT CONFLICT 53 55 /FTID=VAR\_010796.  
 FT CONFLICT 660 660 E -> K (IN MHA).  
 FT CONFLICT 869 869 EAI -> RGH (IN REF. 3).  
 FT CONFLICT 931 931 T -> S (IN REF. 3).  
 FT CONFLICT 931 931 T -> M (IN REF. 4).  
 FT CONFLICT 1240 1241 C -> Y (IN REF. 4).  
 FT CONFLICT 1350 1350 KG -> GR (IN REF. 4).  
 FT CONFLICT 1764 1764 E -> EE (IN REF. 2).  
 FT CONFLICT 1771 1771 T -> A (IN REF. 2).  
 FT CONFLICT 1771 1771 S -> G (IN REF. 2).

Query Match 8.6%; Score 313.5; DB 1; Length 1960;  
 Best Local Similarity 24.0%; Pred. No. 1.2e-06;  
 Matches 150; Conservative 116; Mismatches 273; Indels 87; Gaps 23;

QY 109 AELGKL--LQLVL---GCAISCEKK--QDHQIRIMT-----LEESVQHVWMEAIQELMTK 156  
 DB 971 AKLKKLEEEQIILEDQNCCKLAKEKKLEDDRIAFTNTTEEEKSKSLAKLNKKHEAMIT 1030

QY 157 DTPDSLSPTGVNFDQSRRYVFLSEAEAGD--ELQORCLDLERQLMLLSEKOSLAQE 214  
 DB 1031 DUEERLRE-----EKORQELEKTRRKLEGGSTOLSDIOAELQAO---IALLKMLAKK 1081

QY 215 NAGLRERGRPEGET-PGLTAKKILLQSQLEQENFRLESGREDERLRCALEREV 273  
 DB 1082 EEELQAALARVEEAQAQNMALKKIRELESQISELQD---LESER-----ASRNNK 1129

QY 274 AEQLHRNQAITSLAQAOALKDEMDELROSS-----ERAGOLEATITSCRRRLGELRELR 329  
 DB 1130 AEQKRD-----LGEELKALTELEDTLDSTAQQOELRSKREQEVNILLKKTLEE--EAKT 1182

QY 330 QYRQLEERNAGHAERTROLEDELRRAGSLRAQLEAQRQVQELQORQOEAMKAQKWLFE 389  
 DB 1183 HEAQIOEMRQKHSQAVELEAQOETKRVKANLEKAK---QTLNENRGELANEVVKVLOG 1239

QY 390 CRNLEKYESVTKEKERLLAERDSLREANEEL--RCAQLQPRGLTQADPSLDPTSTPVDN 447  
 DB 1240 KGDSEHKRKKVEAQLOELQVKNEGERVETELADKVKLOVE-LDNVTGLLSQSDSKSK 1298

QY 448 LAAEI--LPAELRETLRLRLQLENKRLCRQEAADRERQEE---LQRHLEDANRARGHLETO 502  
 DB 1299 LTKDFSALESQLODTQELLQEEENRQKLSLTKLKQVEDEKNSFREQLDEEEAKHNLE-- 1356

QY 503 HRLNQOQLSELRAQVEDLQALQEOGGKTEDAISILLKRLKEHLQKLHEADLELQKRE 562  
 DB 1357 -----KOIATLHAQVADMKKMKMDESVGCLETAEEV--KRLQKQDLEGLSORHEEKVAAYD 1409

QY 563 YIEELPPDSDSTARRIEELQH-----NLQKRD-----ADLRAMEERYRYRYVDKARM 609  
 DB 1410 KLEKTYRQLQELDDLVDLDHQRSACNLEKKQKFKDQLLAEKTIKAKYAEERDRAEA 1469

QY 610 VMQTMPEKQORPAAGA-----PPELHSLRTOQLRERDVIRHLEMDPEKFSQRQOEKEL 662  
 DB 1470 EAREKETKALSRLARALEEAAMEQKAELERLNKQFRTMEDMLMSKDDVGKSVHELEKSKRA 1529

QY 663 LISAWYMGMALOORAGERAFAHAQ 688  
 DB 1530 LEQQVEEMKTOLEDELEQATEDAK 1555

RESULT 10  
 MYS\_AEQIR  
 ID MYS\_AEQIR STANDARD; PRT; 1938 AA.  
 AC P24733;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin heavy chain, striated muscle.

OS. Aequipecten irradians (Bay scallop).  
 OC. Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoida;  
 CC. Pectinoida; Pectinidae; Argopecten.  
 OX. NCBI\_TaxID=31199;  
 RN. [1]  
 RP. SEQUENCE FROM N.A.  
 RC. TISSUE=Adductor muscle;  
 RX. MEDLINE=92011595; PubMed=1917970;  
 RA. Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;  
 RT. "Complete primary structure of a scallop striated muscle myosin heavy chain. Sequence comparison with other heavy chains reveals regions that might be critical for regulation.";  
 RT. J. Biol. Chem. 266:18469-18476(1991).  
 RL. [2]  
 RN. SEQUENCE FROM N.A.  
 RC. TISSUE=Adductor muscle;  
 RX. MEDLINE=91083319; PubMed=2263488;  
 RA. Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;  
 RT. "Nucleotide sequence of full length cDNA for a scallop striated muscle myosin heavy chain.";  
 RT. Nucleic Acids Res. 18:7158-7158(1990).  
 RL. [3]  
 RN. X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 777-836.  
 RX. MEDLINE=94173332; PubMed=8127365;  
 RA. Xie X., Harrison D.H., Schlichting I., Sweet R.M., Kalabokis V.N., Szent-Gyorgyi A.G., Cohen C.;  
 RT. "Structure of the regulatory domain of scallop myosin at 2.8-A resolution.";  
 RT. Nature 368:306-312(1994).  
 RL. [4]  
 RN. X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 777-836.  
 RX. MEDLINE=96419133; PubMed=8805510;  
 RA. Houdusse A., Cohen C.;  
 RT. "Structure of the regulatory domain of scallop myosin at 2-A resolution: implications for regulation.";  
 RT. Structure 4:21-32(1996).  
 CC. -1- FUNCTION: MUSCLE CONTRACTION.  
 CC. ACTIVITY: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE ACTIVITY THAT IS ACTIVATED BY F-ACTIN.  
 CC. -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC. -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.  
 CC. -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC. -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
 CC. This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC. EMBL; X55714; CAA39247.1;  
 DR. PIR; S13557; S13557;  
 DR. PIR; A40997; A40997;  
 DR. PDB; 1SCW; 30-APR-94.  
 DR. PDB; 1WDC; 11-JUL-96.  
 DR. InterPro; IPR000048; IQ\_region.  
 DR. InterPro; IPR004009; Myosin\_N.  
 DR. InterPro; IPR002928; Myosin\_tail.  
 DR. InterPro; IPR002017; Spectrin.  
 DR. InterPro; IPR001609; myosin\_head.  
 DR. Pfam; PF00063; myosin\_head; 1.  
 DR. Pfam; PF00612; IQ; 2.  
 DR. Pfam; PF01576; Myosin\_tail; 1.  
 DR. Pfam; PF02736; Myosin\_N; 1.  
 DR. PRINTS; PR00193; MYOSINHEAVY.  
 DR. ProDom; PD000355; myosin\_head; 1.  
 DR. SMART; SM00015; IQ; 1.  
 DR. SMART; SM00242; MYSC; 1.  
 DR. PROSITE; PS50096; IQ; 1.

KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 KW ATP-binding; Alkylation; Calmodulin-binding; 3D-structure.  
 FT DOMAIN 1 777  
 FT DOMAIN 778 805  
 FT DOMAIN 806 836  
 FT DOMAIN 837 868  
 FT NP\_BIND 176 183  
 FT MOD\_RES 693 693  
 FT MOD\_RES 703 703  
 FT HELIX 778 821  
 FT TURN 822 823  
 FT HELIX 825 833  
 FT TURN 834 835  
 SQ SEQUENCE 1938 AA; 222821 MW; ASCCE4127DIA4896 CRC64;

Query Match 8.6%; Score 312; DB 1; Length 1938;  
 Best Local Similarity 20.1%; Pred. No. 1.3e-06;  
 Matches 168; Conservative 153; Mismatches 287; Indels 228; Gaps 30;  
 QY 29 QDLSSGLAVAVLQIDPSWF---NEAWLOGISEDGPGNWL--KVSNLKMWLSLVEYS 83  
 DB 802 QDQIGLS---VIQRTIRKVLVLRNMQW-----WKLYSKVRPLLSIARQEEEMK 847  
 QY 84 QDVLAHPVSEHLPDVSILIGFSPAEGLKLLQLVLGCAISCEKKQDHIQIMTLEESVQ 143  
 DB 848 EQLKQMDKMKEDLAKTERIKKELEQNVTLL-----EQKNDLFLOQLTLEDSMG 896  
 QY 144 HVMEATQELMTKTPDPSLSPETYNFDSQSRYYFLFSEAEDEGLQQRCLDLERLML 203  
 DB 897 D-QEEREKLIQK-----ADFESQIK-----ELEERLLDEDAAD 932  
 QY 204 LSEKQSLAQENAGLRERMRPE-----GEGTGLTAKKLLLLLSQLEQ----- 247  
 DB 933 LEGIKKMEADNANLKDKIGDLENTLQAEQDKAHKQNIQSTLQGEISQDEHIGKLNKE 992  
 QY 248 ---LQEEFNRESGREDELR-----ALEREVAELQHR-----NQALTS 286  
 DB 993 KKALEANKKTSLSQAEDEKCNHNLKAKLEQALDELDNLEREKVRGVDEKARKV 1052  
 QY 287 AQEAQALKDEMDEL---RQSSERAGQLEATLTSCRRRLGE---LRLRRQVQLEERN 338  
 DB 1053 EQDLKSTQENVDELVRKRELEENVRRKEAIISSNLSKLEDEQNLVSQLRKIKELQ---- 1109  
 QY 339 AGHAERTQLEDELRRAGSLRAQLEAQR---ROVQELQQRQEEA-----MKA 383  
 DB 1110 ---ARIEELSEELEAERNARAKVEKQRAELNRELEEL-GERLDEAGGATSAQIELNKKR 1164  
 QY 384 EKWLFEK-RNLEE-----KYESVTKEKERLLAERDSLEA 417  
 DB 1165 EAELKIRDLLEASLQHEAQISALRKKHQDAANEMADVDQLOKVKSKLEKDKDLKRE 1224  
 QY 418 NEELR-----CAQLQPRGLTQADPSLDPTSTPVDNLAAEILPAELRFTLLRLQ 466  
 DB 1225 MDDLSEQTHNMKNKGCSE---KVMKFESQMSDLNARLESQRSI---NELQSKSLQA 1279  
 QY 467 ENKRLCRQ--EAADR-----ERQELQRHLEDANRARGHLETOHRLNQQLSE 512  
 DB 1280 ENSDLTRQLEDAEHRVSVLSKEKSQSLQSDARRSLEETRASKRLQNEVRNMHADMDA 1339  
 QY 513 LRAQVEDLQKA---LQEOGGKTEDAISILLKRLKEHLQKLEADLELQKREYIELEP 569  
 DB 1340 IREQLSEESKSDVQRLSKANNEIQOWRSKFSEGANTEELEDQKRLKGLKLSAEQ 1399  
 QY 570 PTDSSSTAR-----RIEELQHNQKDDADLRAMEERYR-----RYVDKARM 609  
 DB 1400 TTEAANAKCSALEKAKSLRQLEEDMSIEVDNRANASVQMEKKORAFDKTAEQAKVNS 1459  
 QY 610 VMQTMPEKQRPAGAPPPELHSLRTQLRERDVRIRHLEM----- 647  
 DB 1460 LQSELENSQKESRGVSAELVRIKASIEEYQDSIGALRRENKLNLADETHDLTDQSEGRS 1519  
 QY 648 --DFEKSRSQREKEKLLISAWYNGMALQORAGEERAPAHQSFQAQORLATNSR 701

Db 1520 THELDKARRRLEMEKEELOALAEBALEA---EBAKVMRAQLEATVNRNEIDKR 1572

RESULT 11  
MYHA\_HUMAN STANDARD; PRT: 1976 AA.  
AC P35380;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update).  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,  
type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).  
GN MYH10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN [2]  
RP MEDLINE=96025307; PubMed=7499478;  
RX MEDLINE=91316803; PubMed=1860190;  
RX Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,  
Gula B., Adelstein R.S., Weir L.;  
RA "Human nonmuscle myosin heavy chains are encoded by two genes located  
on different chromosomes."  
RT J. Muscle Res. Cell Motil. 16:379-389(1995).  
RL Circ. Res. 69:530-539(1991).  
CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,  
CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND  
CAPING.  
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY  
CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2  
REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M69181; AAA99177.1;  
CC PIR: B61231; B61231.  
CC HSP: P10587; IBR2.  
CC Genew: HGNC:7568; MYH10.  
CC MIM: 160775;  
CC InterPro: IPR000048; IQ\_region.  
CC InterPro: IPR004009; Myosin\_N.  
CC InterPro: IPR002928; Myosin\_tail.  
CC InterPro: IPR002017; Spectrin.  
CC InterPro: IPR001609; myosin\_head.  
CC Pfam: PF00063; myosin\_head; 1.  
CC Pfam: PF00612; IQ; 1.  
CC Pfam: PF01576; Myosin\_tail; 1.  
CC Pfam: PF02736; Myosin\_N; 1.  
CC PRINTS: PR00193; MYOSINHEAVY.  
CC ProDom: PD000355; myosin\_head; 1.  
CC SMART: SM00015; IQ; 1.  
CC SMART: SM00242; MYSC; 1.  
CC PROSITE: PS00936; IQ; 1.  
CC MYOSIN; ATP-binding; Calmodulin-binding; Actin-binding;  
KW Coiled coil; Alkylation; Multigene family.

FT DOMAIN 1 785 MYOSIN HEAD-LIKE.  
FT DOMAIN 786 815 IQ.  
FT DOMAIN 845 1976 COILED COIL (POTENTIAL).  
FT NP\_BIND 178 185 ATP (POTENTIAL).  
FT MOD\_RES 701 711 ALKYLATION (SH-1) (POTENTIAL).  
FT MOD\_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).  
SQ SEQUENCE 1976 AA; 228938 MW; B2BB87FF35EA124F CRC64;  
Query Match 8.4%; Score 306.5; DB 1; Length 1976;  
Best Local Similarity 24.3%; Pred. No. 2.3e-06;  
Matches 169; Conservative 117; Mismatches 285; Indels 125; Gaps 28;  
QY 101 LIGFSD-PAELGKLLQLVLGCAISCEKKQDHIQRTM-----TLSESVQHVMEAIQEL 153  
Db 1060 LDGETTDLQDAELQAQIDELKLAKEEELQGLARGDDTETLKNNAKLVVRELQAQ 1119  
QY 154 MTKDTDPDSLPETYGNFDSQSRYYFLSEAE-GELOQRCLDLERQLMLLSEEQSLA 212  
Db 1120 IAELOEDFESEKASRNKAQKR--DLSELEALKTELEDTLDTAAQQLRTKREQVA 1177  
QY 213 QENAGLRMRGREG--EGTPGLTAKLLLLQSQLEQLEENFLESGR-----EDERLR 265  
Db 1178 ELKKALEETKNEHAQIQDMRQHATALELSQLOAKRFRKANLEKNKOGLETDNKELA 1237  
QY 266 C--AELEREVAELQHRNOALTSQAQAALKDEMDLR-QSSERAGOLEATLTSCRRRLG 322  
Db 1238 CEVVLQOVKAESHRKKLDAQVQELHAKVSGDRLRVELAELAKSKLQNELDNVSTLE 1297  
QY 323 ELRE-----LRRQVRLQEE-----RNAGHAETROLEDELRAGSLRAQLEA 364  
Db 1298 EAEKGIKFAKDAASLESQDQLQELQETROKLNLSRIQLEEE---KNSIQEQEE 1354  
QY 365 Q-----RQVLOEQROEEAMK-----AEKWLFE-----CRNLEEK--- 396  
Db 1355 EEEARKNLEQVLAQSQALADTKKKVDDDLGTIESLEAKKKLLKDAELASQREELKALA 1414  
QY 397 YESVTKEKERLLAERDSI-----REANEELRCAQLQPRGLTQADPSLDPT-STPYDNL 448  
Db 1415 YDLEKTKNRILQQLDLDLTDVLDHQRQVANSLEKQKKFDOLLAEEKSISARYAERDRA 1474  
QY 449 AAETLPALRETLRLQLENKRLCQEAADPERQ-----EELQRHLEDANRARGLETQH 503  
Db 1475 EAEARETKALSARALEEALEAKEEPERQKQLRADMEDLMSSKDDVGNVHELEKSK 1534  
QY 504 RLNQOQSELRAQVEDLOKALQEOGGKTEDAISILLKKLEEHQKLH-EADLELQKRE 562  
Db 1535 RALQOQVEEMRTQLELEDELEQ-----ATEDA-----KURLEVNQMAKQAFERDQTRDE 1585  
QY 563 YIELEPPTDSSTARRIEELQHNKQKDALRAMEERYRYV-----DKARNVMQTMPEPK 617  
Db 1586 QNEKK-----RLLIKQVRELEALE-----DERQALAVASKKKKMEIDKDLAQ 1632  
QY 618 QPAAQAPPELHSLRTRERDVRIRHLEMDFEKRSORE-----QEEKLLISAWYNNM 670  
Db 1633 IEAANKARDEVIK---QLRKLQAQMKQYQRELEEARARSDIEFAQSKSEKKLKLEABI 1689  
QY 671 GMAQOQAGEERAPAHQAQSLAQQLA---TNSRRG 703  
Db 1690 LQLOEELASSERARRHAQ---ERDELADEITNSAG 1723

RESULT 12  
RA50\_ARCFU  
ID RA50\_ARCFU STANDARD; PRT: 886 AA.  
AC O29230;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA double-strand break repair rad50 Atpase.  
GN RAD50 OR AF1032.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.







RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,  
 RA Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:31-39(1998).  
 CC -1- FUNCTION: COMPONENT OF THE CORNIFIED ENVELOPE OF KERATINOCYTES.  
 CC MAY LINK THE CORNIFIED ENVELOPE TO DESMOSOMES AND INTERMEDIATE  
 CC FILAMENTS.  
 CC -1- SUBUNIT: HOMODIMER OR A HETERODIMER WITH EVPL (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH DESMOSOMES AND INTERMEDIATE  
 CC FILAMENTS.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN STRATIFIED SQUAMOUS EPITHELIA AND  
 CC IN SOME OTHER EPITHELIA.  
 CC -1- INDUCTION: DURING DIFFERENTIATION OF EPIDERMAL KERATINOCYTES.  
 CC -1- SIMILARITY: CONTAINS 2 PLECTIN REPEATS.  
 CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE PLECTIN OR CYTOLINKER FAMILY.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF001691; AAC17738.1;  
 CC EMBL: AF013711; AAC39668.1;  
 CC EMBL: AF041004; RAD17459.1;  
 CC EMBL: AF040999; RAD17459.1; JOINED.  
 CC EMBL: AF041000; RAD17459.1; JOINED.  
 CC EMBL: AF041002; RAD17459.1; JOINED.  
 CC EMBL: AF041003; RAD17459.1; JOINED.  
 CC EMBL: AB011140; BAA25494.1;  
 CC Genbank: HGNC:9273; PPL.  
 CC MIM: 602871;  
 CC InterPro: IPR001101; Plectrin\_repeat.  
 CC InterPro: IPR002017; Spectrin.  
 CC Pfam: PF00681; Plectrin; 1.  
 CC SMART: SM00250; PLEC; 2.  
 CC Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein.  
 CC -----  
 CC FT DOMAIN 16 125  
 CC FT REPEAT 188 389  
 CC FT REPEAT 216 317  
 CC FT REPEAT 323 485  
 CC FT REPEAT 505 612  
 CC FT REPEAT 733 861  
 CC FT DOMAIN 585 820  
 CC FT REPEAT 886 1645  
 CC FT REPEAT 1651 1685  
 CC FT REPEAT 1700 1735  
 CC FT CONFLICT 589 589  
 CC FT CONFLICT 657 657  
 CC FT CONFLICT 994 994  
 CC FT CONFLICT 1573 1573  
 CC FT CONFLICT 1663 1663  
 CC SEQUENCE 1756 AA; 204649 MW; 557C7D92BE18F107 CRC64;  
 Query Match  
 Best Local Similarity 21.4%; Score 304.5; DB 1; Length 1756;  
 Matches 183; Conservative 129; Mismatches 270; Indels 275; Gaps 33;  
 QY 68 KYSNKLWLSLVEYSQDVLAHPVSEHLPDVSGLICFSDPAELGKLLQVLGCAISCEK 127  
 DB 856 RLQNLFEALNLRQQEVEVTHETLQNRPDSPGVSEAWKIRKELDE-----ETER 905  
 QY 128 KODHIQIMTLESVOHYVNEAQELMT-----KDPDLSLPTETGNYFSDSRRYFLSE 182  
 DB 906 RROLENEVSTQBEIWTLRNQGQSVVRKEVLKVPDPVLESF-----QQQR--TLAE 959  
 QY 183 EABEGDELAQRCILDLERQLMLLSEERQ-----SLAENAG 217

DB 960 EQHKNLQLEFEALQQLRALRQETRDGQGVVYKRVIEIPDRAQAEVLQREELEA 1019  
 QY 218 LRERMGPRPE-----GTPGLTAKLLLLQS--QLE-----QLQENR 254  
 DB 1020 LRQKGAREAEVLLLOORVAALAEKSRQAEKVEYVKVQLNDPQLEAEYQOQEDHOR 1079  
 QY 255 LESGRE-----DERLRCAELEVEAELOHRNALTSLAQEAQALKDEMDELRSSE 305  
 DB 1080 QDOLRKEQEEELSFLQDKLKRLEKRAMAGKITVVEVLKVEKDAATERVSDLTROYED 1139  
 QY 306 RAGQLEATITSCRRRLGELRELRRVQLEERNAGHA--ERTRL-----348  
 DB 1140 EAAKARASOR-----ETELLRLKIWALEENAKVVQVEKVIYRPPDKAESEVANLR 1192  
 QY 349 -----EDELRRAGSLRAQLEARR--QVQ-----ELQ 373  
 DB 1193 LEIVQERKYRGAEQLR---SYQSELEALRRRGPOVQVEKVEIKYKTDPEMEKELO 1249  
 QY 374 GQEQE-----EAMKAEKWLFEGRNL-----393  
 DB 1250 RLREEIVDKTRLIERCDLEIYQLKKEIQALKDTKPOVQPKVVOETLQEQDPQTKEEVA 1309  
 QY 394 -----EKYSVTVKEKRELLAERDSLRANBELRCALQPRGLTQ-----ADP-----436  
 DB 1310 SLRAKLSEEQKKQVDLERER-ASQEQIARKEEL--SRVKERVQGVVYVEEPEGLRA 1366  
 QY 437 -----SLDPTSTPVDNLAAETLPAELRETLLRLQLENKRLICRQEAADRERQELQR-- 487  
 DB 1367 EASAFAESIDVELRQIDKLRAELRLRLRRRTLELERQLE--ELERERQARREAREVORLQ 1424  
 QY 488 -----HL-----EDANRAR-HGL-----ETQHR--LQQQQLSELR 514  
 DB 1425 QRLAALQEAEAEAREKVTHTQKVVLOQDPQAREHALLRLQLEEEQHRRLLEGETLIR 1484  
 QY 515 AQVEDLQKA-----LQEQGGKTEDAISILKRLKEELHQLKLEADLELQKREY 563  
 DB 1485 RLAALEKAEVKEKVVLSVQVQVKGDTQEQIQ-RLKSSLEESRSKRLDVEVSRLER 1543  
 QY 564 IELEPPTDSSARR--IEELQHNLOKKDADLRAMEERYRYVDKARMYMTMEPKQPA 621  
 DB 1544 LSELEPHNSKSKELDFLAENHKLQLEQNQLQLETRRLQSEINMA--ATETRDLRNMV 1601  
 QY 622 AGA-----PPELSLRTQLRERDVRTHLEMDSEKRSQEQEKLISAWYNM 670  
 DB 1602 ADSTGNHDSRLWSLERELDDLRSLKSKDKLEIDEQLKRLGSAVKREQRENHLRSI---- 1658  
 QY 671 GMAQQORAGEERAPAHA 687  
 DB 1659 -VVIHPDTGRELSPERA 1674  
 RESULT 15  
 RA50\_AQUAE STANDARD; PRT; 978 AA.  
 AC O67124;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Probable DNA double-strand break repair rad50 ATPase.  
 GN RAD50 OR AQ.1006.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
 OC Aquifex.  
 OX NCBI\_TaxID-63363;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RX MEDLINE-98196666; PubMed-9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus.";



RL Nature 392:353-358(1998).

CC -I- FUNCTION: Involved in DNA double-strand break repair (DSBR). The

CC rad50/mrell complex possesses single-strand endonuclease activity

CC and ATP-dependent double-strand-specific exonuclease activity.

CC Rad50 provides an ATP-dependent control of mrell by unwinding

CC and/or repositioning DNA ends into the mrell active site (By

CC similarity).

CC -I- SUBUNIT: Forms a complex with mrell (By similarity).

CC -I- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AAC00718; AAC07092.1; -

CC InterPro; IPR003439; ABC\_transportr.

CC InterPro; IPR004592; SbcC.

CC InterPro; IPR002017; Spectrin.

CC TIGRFAMS; TIGR00618; sbcc; 1

CC DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.

CC NP\_BIND 32 39 ATP (By similarity).

CC FT DOMAIN 160 826 COILED COIL (POTENTIAL).

CC SEQUENCE 978 AA; 115897 MW; 9B0F2BF51ADD1151 CRC64;

CC -----

CC Query Match 8.3%; Score 302; DB 1; Length 978;

CC Best Local Similarity 22.1%; Pred. No. 1.8e-06;

CC Matches 147; Conservative 122; Mismatches 220; Indels 176; Gaps 25;

CC -----

CC QY 65 WKLYSNL-----KWLRLSVYSQDVLHPVSEHLPDVSIGEF-----SDPAELQKL 114

CC DB 120-WLFKISGLDYKFTKVL-----LPQ-----GEFDFLKESSEKKI 156

CC QY 115 LQVLGLCAISCEKKQDHQIRIMTLESQHVVNEAIQELMTKDTPDSLSPTYGNFDSQ- 173

CC DB 157 LINLLG-----LEE-----LEKVRQ-----LASETFKNLEGR 184

CC QY 174 ----SRYYFLSEAEGBDELQQRCLDLERQLMLLSEKQSLAQENAGLRMRGRPEGET 230

CC DB 185 EALKKEYELLDYPTPKVE-----LEKTLNLEBELKELTEKTELQELKKAEEKDS 238

CC QY 231 PGLTAKKLLQSLQLEENFRLESGREDE-----LRCAELREYVAELQHRNQ 281

CC DB 239 LERELSQVTKLELENLEKEVEKLEKLEFSKVPYPIAKRIEIDKLTLEKVRKN 298

CC QY 282 ALTSLAQEAQALDEM-----DELROSERAGQLEATLTSCRRRLGELREL 327

CC DB 299 KLT---KELAVLKDELQSLQLEENFRLESGREDE-----LRCAELREYVAELQHRNQ 348

CC QY 328 RQVRLQLEERNAGHAEKTRQLEDELRAQSLRAQLEAQRQVQLEQORQOEAMKAQKWL 387

CC DB 349 KEILKELSSLSKEREYQAKQEFEDLSERVEKQKLVAE-----TEELEKIKEL 403

CC QY 388 FECLNLEKVEYVTKERLLAERDLSREANEELRAQQLQPRGLTQADPSLDPTSTPVDN 447

CC DB 404 FS-----EEETSL-KKERLLVE---LQRLKELKEGEGLENLTQYKEKKVHEKVLN 455

CC QY 448 LAABILPAELRETLRLQ-----LENNKLCROEAADRERQBELQR 487

CC DB 456 ELKE-LERELKERELHYAHMVASYLSPGDTCPVGGGIYRGKAL---ENVDAEGISELKH 511

CC QY 488 HLEDANRARGHLETOHRLNQQLSELRAQVEDLQKALQEOGGKTEDAISILLKRLKEHL 547

CC DB 512 AKELKEEREIDTTLKYAKINSKAEEMKLENEVE-----LRKEIPENLAKERI 563

CC QY 548 QKLHEADLE-----LQKREYEELEPPTSDSTARRIELOHNLQKDKADLRAMEERY 600

CC DB 564 KKLBELEKEKEKLEHKLKLYKKALEDNQ-----KQKEAQAQAKHQAQTELELLAKKI 615

CC QY 601 RRYVDKARNV-----MQTMPEKQPAAGAPPHELRLTQLRERDVRIRHLEMOFEKRSQ 655

Search completed: March 4, 2003, 14:59:59

Job time : 36.2109 secs

Db 616 R---EKSLVKPEKELYRVERLEDYEEESLKEEYINSLQIEEKEKKLRKHFEELSSR 672

QY 656 REQEE 660

Db 673 KSKLE 677



Db 203 EFPDEQLRRRLLELRKRGKEEQOORRRQDRVFEKEEKWKRRKRETVLRKEEKLQ 262  
 QY 157 DTPD---SLSPETYNFDSQSRYY--YFLSEAEAGEDELOQRCCLDLERQMLLSEKQSL 211  
 Db 263 EEPQORQLOE---EQLRKLEQLERREERQEEQOOR-LRREQRLRKEEERRE 316  
 QY 212 AENAGLRMRGREGGTPGLTAKKLLQLSQLOEQLENFLESREDELRCAELER 271  
 Db 317 QOER---REQOERRE---QOERREQOER---REQOERREQOER 353  
 QY 272 EVAELOHNRQALTSIAQE---AQALKDEMDLRQSSERAGOLEATLTSRRRLGELR--- 325  
 Db 354 EQEE-ERREQOERREQOERREQOERREQOERREQOERREQOERREQOERREQOER 412  
 QY 326 ELRRQVQLERNAHGAERTQLEDELRRAGSLRAQLAQRQVQOELQOQOEEAMKA 385  
 Db 413 QLARE-QOLRREQOERREQOERREQOERREQOERREQOERREQOERREQOER 470  
 QY 386 ---WLFECRNLEEKYESVTKEKELRLAERDSLREANEELRCAQLQPRGLTQADPSLDP 440  
 Db 471 EERDNL---KREETERHEQERKQOOLKRDQOEEERERWKLKEEERREQO 519  
 QY 441 TSTPVDNLAAEILPAELRETLLRLQLENKRLCRQEAADRERQOELQHLLEDANRARGLE 500  
 Db 520 ---ERREQOERREQOERREQOERREQOERREQOERREQOERREQOERREQOER 560  
 QY 501 TOHRLNQ--QOLSELRAQVEDLQALQOEGGKTEDAISILLKRLKEHLQKL---HEADL 555  
 Db 561 QERLEQLKREBEKLEQERREORLKRQOERRDQ---LLKREERROQLRKEOERL 617  
 QY 556 ELQKREYIELEP-----PTDSSTARRIEELQHLNQ----- 587  
 Db 618 EQLKREEVERLEQOERREDELRKREPEERREHLLKSEOEERREQOERREQOER 677  
 QY 588 -KKDADLRAMEERYRYVDKARVMQTMPEKQPA-----AGAPPELHSLRQLRDRVR 641  
 Db 678 LKREEEERLEORLKRHEEERREQOERREQOERREQOERREQOERREQOERREQOER 737  
 QY 642 IRLHLEMDFKRSORQOEEKLLISA---WYNGMALQORAGERA-----PAHAQSL 691  
 Db 738 VLLEAPQAGRAEAPQOEEKRRRESELOQWQEEERAHQOEEERDFTWQWQAEKSER 797  
 QY 692 AQORLATNSRRGPLGRLASLNLRTDK 718  
 Db 798 GRQL---SARPPLEQRERQLRAER 821

## RESULT 2

US-08-800-644-94  
 ; Sequence 94, Application US/08800644  
 ; Patent No. 5958752  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steinert, Peter M.  
 ; APPLICANT: Lee, Seung-Chul  
 ; APPLICANT: Kim, In-Gyu  
 ; APPLICANT: Chung, Soo-Il  
 ; APPLICANT: Park, Sang-Chul  
 ; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
 ; NUMBER OF SEQUENCES: 117  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 680 Newport Center Drive, Sixteenth Floor  
 ; CITY: Newport Beach  
 ; STATE: CA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/800,644  
 ; FILING DATE: 14-FEB-1997  
 ; CLASSIFICATION: 424  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/056,200  
 ; FILING DATE: 30-APR-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fedrick, Michael F.  
 ; REGISTRATION NUMBER: 36,799  
 ; REFERENCE/DOCKET NUMBER: NIH054.001A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (714) 760-0404  
 ; TELEFAX: (714) 760-9502  
 ; INFORMATION FOR SEQ ID NO: 94:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1898 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-800-644-94

## Query Match

Best Local Similarity: 8.7%; Score 315; DB 2; Length 1898;  
 Matches 168; Conservative 110; Mismatches 269; Indels 140; Gaps 29;

QY 104 EFSDPALCK--LLQI-VLGCRAISCEKQDHQIQRIMTLESVQ-----HYVMEAIQELMTK 156  
 Db 203 EFPDEQLRRRLLELRKRGKEEQOORRRQDRVFEKEEKWKRRKRETVLRKEEKLQ 262  
 QY 157 DTPD---SLSPETYNFDSQSRYY--YFLSEAEAGEDELOQRCCLDLERQMLLSEKQSL 211  
 Db 263 EEPQORQLOE---EQLRKLEQLERREERQEEQOOR-LRREQRLRKEEERRE 316  
 QY 212 AENAGLRMRGREGGTPGLTAKKLLQLSQLOEQLENFLESREDELRCAELER 271  
 Db 317 QOER---REQOERRE---QOERREQOER---REQOERREQOER 353  
 QY 272 EVAELOHNRQALTSIAQE---AQALKDEMDLRQSSERAGOLEATLTSRRRLGELR--- 325  
 Db 354 EQEE-ERREQOERREQOERREQOERREQOERREQOERREQOERREQOERREQOER 412  
 QY 326 ELRRQVQLERNAHGAERTQLEDELRRAGSLRAQLAQRQVQOELQOQOEEAMKA 385  
 Db 413 QLARE-QOLRREQOERREQOERREQOERREQOERREQOERREQOERREQOER 470  
 QY 386 ---WLFECRNLEEKYESVTKEKELRLAERDSLREANEELRCAQLQPRGLTQADPSLDP 440  
 Db 471 EERDNL---KREETERHEQERKQOOLKRDQOEEERERWKLKEEERREQO 519  
 QY 441 TSTPVDNLAAEILPAELRETLLRLQLENKRLCRQEAADRERQOELQHLLEDANRARGLE 500  
 Db 520 ---ERREQOERREQOERREQOERREQOERREQOERREQOERREQOERREQOER 560  
 QY 501 TOHRLNQ--QOLSELRAQVEDLQALQOEGGKTEDAISILLKRLKEHLQKL---HEADL 555  
 Db 561 QERLEQLKREBEKLEQERREORLKRQOERRDQ---LLKREERROQLRKEOERL 617  
 QY 556 ELQKREYIELEP-----PTDSSTARRIEELQHLNQ----- 587  
 Db 618 EQLKREEVERLEQOERREDELRKREPEERREHLLKSEOEERREQOERREQOER 677  
 QY 588 -KKDADLRAMEERYRYVDKARVMQTMPEKQPA-----AGAPPELHSLRQLRDRVR 641  
 Db 678 LKREEEERLEORLKRHEEERREQOERREQOERREQOERREQOERREQOERREQOER 737  
 QY 642 IRLHLEMDFKRSORQOEEKLLISA---WYNGMALQORAGERA-----PAHAQSL 691  
 Db 738 VLLEAPQAGRAEAPQOEEKRRRESELOQWQEEERAHQOEEERDFTWQWQAEKSER 797  
 QY 692 AQORLATNSRRGPLGRLASLNLRTDK 718  
 Db 798 GRQL---SARPPLEQRERQLRAER 821

RESULT 3  
US-08-466-390-4  
; Sequence 4, Application US/08466390  
; Patent No. 5686562  
; GENERAL INFORMATION:  
; APPLICANT: TOURKATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,390  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESO, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-466-390-4

RESULT 4  
US-08-470-950-4  
; Sequence 4, Application US/08470950  
; Patent No. 5698439  
; GENERAL INFORMATION:  
; APPLICANT: TOURKATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,950  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESO, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

Query Match 8.6%; Score 311.5; DB 1; Length 2101;  
Best Local Similarity 23.2%; Pred. No. 2.8e-15;  
Matches 198; Conservative 130; Mismatches 294; Indels 233; Gaps 36;  
QY 4 DKALCGSLTLWQTHVPSPCAS--PQDLSSGLAVA-----YVLNQIDPSWFNEAWLQ 55  
DB 139 DGLNLNEDLENFLQAPVPTCSSTFPPELSPHQAKREIRFLEQLQKVASSSSGNNFLS 198  
QY 56 GISEDGPNWKLKVNKLWLSLVEYSQDVLAPVSEHLPDVSILGFSFDPALGKLL 115  
DB 199 GSPASP-----MGDILOTQFQWRRLKKQLADRSNRDELE--L 235  
QY 116 QLVLAGCAISCEK-----KQDHQIRIMTLEESVQHVYME--AIQELMTKDTDPDSLSPETY 167  
DB 236 ELAENRKLKTEKDAQIANMQRIDRLALNENKQAASPLEPKELEEL--RDKNESLTMRLH 293  
QY 168 GNF-----DSQSRYYFLSEAEED-----ELQQRCLDLEROLMLLSEKQSL 211  
DB 294 ETLKQCQDLKTEKSOMDRKINOLSE--ENGDSFKLREFASHLQQLQDALNELTEHNSKA 351  
QY 212 AQENAGLRMRGPEGPGTAKLL-----LLQSQLEQLQEEFNRESGREDEER--- 263  
DB 352 TQE---WLEKQKLEKLSAALQDKKLEKNEKNEILQKLSLEELHLSQQLQDNPPEKGEV 408  
QY 264 -----LRCAELEREVALQHRN-----QALTSLSA 287  
DB 409 LGDVLEQLTKQENATLANNTQLOARVEMLETERCQEQAKLLAERGHFEERKEQQLSSLI 468  
QY 288 QEAQALKDEMDLRQSSERAGOLE-----ATLTSRRRLGELRELRRQVRQLEERNA 339  
DB 469 TDLOSSISNLQSAKELEQASQAHGARLTAQVASTS-----ELTLNATIQQQQQLA 522

US-08-470-950-4

Query Match 8.6%; Score 311.5; DB 1; Length 2101;  
Best Local Similarity 23.2%; Pred. No. 2.8e-15;  
Matches 198; Conservative 130; Mismatches 294; Indels 233; Gaps 36;

QY 4 DKAELCGSLTLWQTHFVSPCAS--PDLSSGLAVA-----YVNLQIDPSWFNEAWLQ 55  
DB 139 DGLNEDNEFLQKAPVSTCSSTFFPEELSPSHQAKREIRFLQKVASSSSGNNFLS 198  
QY 56 GISEDGPNKVLKVSNLKWLRSVSDVLAHPVSEHLPDVSLIGFSDPAELGKLL 115  
DB 199 GSPASP-----MGDILQTPQFMRLKKQLADERSNRDELE--L 235  
QY 116 OLVLGCAISCEK-----KODHIOIRMTLESVQHVME--AIQELMTKDPDLSLSPETY 167  
DB 236 ELAENRKLITKDAQIAMMQORIDRLALLNEKQAAAPLEPKELEEL--RDNESLTMRLH 293  
QY 168 GNF-----DSQSRYYFLSEAEED-----ELQORCLDLERQLMLLSEKQSL 211  
DB 294 ETLKOCQDLKTEKSQMDRKINQISE--ENGDLSPKREFASHLQOLQDALNELTEHKA 351  
QY 212 AOENAGLRMRGPEGTGTLAKLL-----LQSOLEQLOEENFLESGRDER--- 263  
DB 352 TQE---WLEKQAKLEKLSAALQDKCKLEEKNEILOGKLSOLEHLSLOLQNDPPOKEGEV 408  
QY 264 ---LRCAELEREVAELQHRN-----LRCAELEREVAELQHRN-----QALTSIA 287  
DB 409 LGDVLQLETLKQEAATLAANNTOQARVEMLETERGQOQAKLLAERGHFEKQOQLSLI 468  
QY 288 QEAQALKDEMDELROSSERAGOLE-----ATLSCRRRLGELRELROVROLEERNA 339  
DB 469 TDLOSSISNLSQAELEEQASQAHARITAOVASLTS-----ELTLNATIQOODOLA 522  
QY 340 GHAERTRQLEDELRRAGSLRAQLEAQ---RRQVQELQ---ORQEEAMK--AEKWLFEERN 392  
DB 523 GLKQAKKEKQAKL--AOTLQOQOASQGLRHQVQELSSSLKQKQEQQLKEVAEKQEAATQD 580  
QY 393 LEEKYESVTKERLLAERDSL---REANEELRCAQLQ-----PRGLTQ 433  
DB 591 HAQOLATAAEERASLRERDAALKOLEALEKEKAALILELOQOQLOVANEARDSAQTSTVQ 640  
QY 434 ADPSLDPTSTPDVNLAA-----EILPAELRETLRLQLENKRLCROEAADRER----- 481  
DB 641 AQREKAELSRVVEELQACVETARQEQHEAQVAELQELQISE---QKATEKERVQAEK 697  
QY 482 ---QELQRLHEDNANRHLETOHR-----LNOQQ---LSELRAQVEDL-----QKAL 524  
DB 698 DLOEQLOALKESLKVTKGSEEEKRRAADALEEQORCISELKAEATRSVLVQHKRERKEL 757  
QY 525 QE-----OGKTEDAISILKRLKLEHLQKLKLEHDLQKREYIEELEPPT 571  
DB 758 EERAGRKGLEARLLQGEAHOAEYEVRLRELAEMAAQHTAESECE---OLVKEVAWR 814  
QY 572 DSSTARTRIEELQHNKQKDALRAMEERYRYVDKARMVMTMEPKORPAA----- 622  
DB 815 DGYEDSQEEBAQYAMFQELMTKEE-----CEKARQELQ--EAKKQVAGIESHSELQI 867  
QY 623 ---GAPPELHS---LRTQLRERDVIRHLEDF-----EKSRSQOREQEKLLISAWNM 670  
DB 868 SRQONKLAELHANLALQOQVEKVRQAQLADDLSTLOEKMAATSKEVAL----- 919  
QY 671 GMALOQRAGEERAPA 685  
DB 920 -ETLVKAGEQOETA 933

RESULT 5

US-08-467-781-4

; Sequence 4, Application US/08467781

; Patent No. 5780596

; GENERAL INFORMATION:

; APPLICANT: TOURATLY, GARY

APPLICANT: LIDGARD, GRAHAM P  
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: 125 HIGH STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,781  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER ESO, EDMUND R  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: MTP-013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2101 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-467-781-4

Query Match 8.6%; Score 311.5; DB 1; Length 2101;  
Best Local Similarity 23.2%; Pred. No. 2.8e-15;  
Matches 198; Conservative 130; Mismatches 294; Indels 233; Gaps 36;

QY 4 DKAELCGSLTLWQTHFVSPCAS--PDLSSGLAVA-----YVNLQIDPSWFNEAWLQ 55  
DB 139 DGLNEDNEFLQKAPVSTCSSTFFPEELSPSHQAKREIRFLQKVASSSSGNNFLS 198  
QY 56 GISEDGPNKVLKVSNLKWLRSVSDVLAHPVSEHLPDVSLIGFSDPAELGKLL 115  
DB 199 GSPASP-----MGDILQTPQFMRLKKQLADERSNRDELE--L 235  
QY 116 OLVLGCAISCEK-----KODHIOIRMTLESVQHVME--AIQELMTKDPDLSLSPETY 167  
DB 236 ELAENRKLITKDAQIAMMQORIDRLALLNEKQAAAPLEPKELEEL--RDNESLTMRLH 293  
QY 168 GNF-----DSQSRYYFLSEAEED-----ELQORCLDLERQLMLLSEKQSL 211  
DB 294 ETLKOCQDLKTEKSQMDRKINQISE--ENGDLSPKREFASHLQOLQDALNELTEHKA 351  
QY 212 AOENAGLRMRGPEGTGTLAKLL-----LQSOLEQLOEENFLESGRDER--- 263  
DB 352 TQE---WLEKQAKLEKLSAALQDKCKLEEKNEILOGKLSOLEHLSLOLQNDPPOKEGEV 408  
QY 264 ---LRCAELEREVAELQHRN-----LRCAELEREVAELQHRN-----QALTSIA 287  
DB 409 LGDVLQLETLKQEAATLAANNTOQARVEMLETERGQOQAKLLAERGHFEKQOQLSLI 468  
QY 288 QEAQALKDEMDELROSSERAGOLE-----ATLSCRRRLGELRELROVROLEERNA 339  
DB 469 TDLOSSISNLSQAELEEQASQAHARITAOVASLTS-----ELTLNATIQOODOLA 522  
QY 340 GHAERTRQLEDELRRAGSLRAQLEAQ---RRQVQELQ---ORQEEAMK--AEKWLFEERN 392  
DB 523 GLKQAKKEKQAKL--AOTLQOQOASQGLRHQVQELSSSLKQKQEQQLKEVAEKQEAATQD 580  
QY 393 LEEKYESVTKERLLAERDSL---REANEELRCAQLQ-----PRGLTQ 433

581. HAQALATAAEERASLRDAALQKLEAKLEAKLEILQOQLOVANEARDSAQTSTQ 640  
434 ADPSLDTPSPVDNLA-----EILPAELRETLRLQLENKRLCROEADRR----- 481  
641 AQREKAELSRKVELOACVETARQOEHAQAQVAELQLRSE---QKATEKERVAAEK 697  
482 ---QEELORHLEDANRARGHLETOHR-----LNQOQ--LSELRAQVEDL-----OKAL 524  
698 DQLOEQLOALKESLKVTKGSLSEEEKRRAADALEEQRCISELKAEATSLVQHQKREKEL 757  
525 QE-----OGKTEDAISILLKRLKEHLQKLEHLELQKREYIEELEPPT 571  
758 EEBRAGRKGLEARLLQLOGEAQAETEVLRRELAEMAAQHTAESECE---QLVKEVAWR 814  
572 DSSTARIEELQHNLOKDDADLRAMEERYRYVDKARVMQTMPEKORPAA----- 622  
815 DGYEDSQOEEAQYGMFQEQMLTKEE-----CEKAROELQ--EAKKVGAGIESHSELQI 867  
623 ---GAPPELHS---LRTOLRERDVIRHLEMDF-----EKSRSQROEEKLLISAWNM 670  
868 SRQONKLAELHANLARALQOQVEKEVRAQKLADDLSTLQEKMAATSKVARL----- 919  
571 GMAQQORAGEERAPA 685  
920 -ETLVRKAGEQOETA 933

RESULT 6  
US-08-195-487-4  
Sequence 4, Application US/08195487  
Patent No. 5783403  
GENERAL INFORMATION:  
APPLICANT: TONKATLY, GARY  
APPLICANT: LIDGARD, GRAHAM P  
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA HURWITZ & THIBEAULT  
STREET: 53 STATE STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/195,487  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/901,701  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER ESQ, EDMUND R  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: MTP-013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: #617/248-7000  
TELEFAX: 617/248-7100  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2101 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-195-487-4  
Query Match 8.68; Score 311.5; DB 1; Length 2101;  
Best Local Similarity 23.2%; Pred. No. 2.8e-15;

Matches 198; Conservative 130; Mismatches 294; Indels 233; Gaps 36;  
QY 4 DKALCGLTWTQTFHVPSPCAS--PQDLSGLAVA-----YVLNQIDPSWFWNEAWLQ 55  
DB 139 DGLNLNEDLENFLOKAPVPSTCSFTFPEELSPSPHQAKREITREFLEOKVASSSSGNNFLS 198  
QY 56 GISDEDGPNKKLVNSLKMVLRLSVLEYSDQVLAHPVSEEHLPDVSLLIGESDPALAEKLL 115  
DB 199 GSPASP-----MGDIQTPQPMRRRLKKQLADERSNRDELE--L 235  
QY 116 QLVGCAISCEK-----KQDHQRIQIMTLEESVGHVME--AIQELMTKQTPDLSLSPETY 167  
DB 236 ELAENRKLKTEKDAQIAMQORIDRLALLNEKQASPLEKPELEEL--RKNESUTMRHL 293  
QY 168 GNF-----DSQSRYYFLSEAEEDG-----ELQORCLDLERQLMLSEKQSL 211  
DB 294 ETLKQCQDLKTEKSDMRKINQNSE--ENGDSLFPKREFASHLQLOQLODALNELTEHESKA 351  
QY 212 AQENAGRLRERMRGEGGTGCLTAKLL-----LQSOLELOQOEENFRLESGREDE-- 263  
DB 352 TOE---WLEKQAQLEKELSAALQDKCLEEKNEILOQKLSOLEEHLQLOQDNPPQEKGEV 408  
QY 264 ---LRCAELEREVAELOHRN-----QALFSLA 287  
DB 409 LGDVLQLETLKQEAATLAANNTOQARVEMLETERGQOEAKLLAERGHFEKQQLSSLI 468  
QY 288 QEAQALKDEWDELQSSERAGOLE-----ATLTCRRRLGELRELRLRQVROLEBERNA 339  
DB 469 TDLOSSISNLSQAKKEELQASQAHGARLTAQVASTS-----ELTTINATIQOQOELA 522  
QY 340 GHAERTQLEDELRRAGSLRAQLEAQ---RQVOELOG--QROEAMK--AEKWLFCERN 392  
DB 523 GLKQAKKEKQAL--AOTLOQOEQASQGLRHQVEQLSSLSKQKQOQLKEVAEKQOEAATROD 580  
QY 393 LEEKYESTVKEKERLLAERDSL---REANBELRCAQLOQ-----PRGLTQ 433  
DB 581 HAQALATAAEERASLRDAALQKLEAKLEAKLEILQOQLOVANEARDSAQTSTQ 640  
QY 434 ADPSLDTPSPVDNLA-----EILPAELRETLRLQLENKRLCROEADRR----- 481  
DB 641 AQREKAELSRKVELOACVETARQOEHAQAQVAELQLRSE---QKATEKERVAAEK 697  
QY 482 ---QEELORHLEDANRARGHLETOHR-----LNQOQ--LSELRAQVEDL-----OKAL 524  
DB 698 DQLOEQLOALKESLKVTKGSLSEEEKRRAADALEEQRCISELKAEATSLVQHQKREKEL 757  
QY 525 QE-----OGKTEDAISILLKRLKEHLQKLEHLELQKREYIEELEPPT 571  
DB 758 EEBRAGRKGLEARLLQLOGEAQAETEVLRRELAEMAAQHTAESECE---QLVKEVAWR 814  
QY 572 DSSTARIEELQHNLOKDDADLRAMEERYRYVDKARVMQTMPEKORPAA----- 622  
DB 815 DGYEDSQOEEAQYGMFQEQMLTKEE-----CEKAROELQ--EAKKVGAGIESHSELQI 867  
QY 623 ---GAPPELHS---LRTOLRERDVIRHLEMDF-----EKSRSQROEEKLLISAWNM 670  
DB 868 SRQONKLAELHANLARALQOQVEKEVRAQKLADDLSTLQEKMAATSKVARL----- 919  
QY 671 GMAQQORAGEERAPA 685  
DB 920 -ETLVRKAGEQOETA 933

RESULT 7  
US-08-483-924-4  
Sequence 4, Application US/08483924  
Patent No. 5882876  
GENERAL INFORMATION:  
APPLICANT: TONKATLY, GARY  
APPLICANT: LIDGARD, GRAHAM P  
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
 STREET: 125 HIGH STREET  
 CITY: BOSTON  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/483,924  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: PITCHER ESQ, EDWARD R  
 REGISTRATION NUMBER: 27,829  
 REFERENCE/DOCKET NUMBER: MTP-013  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 248-7000  
 TELEFAX: (617) 248-7100  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2101 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-483-924-4

Query Match 8.6%; Score 311.5; DB 2; Length 2101;  
 Best Local Similarity 23.2%; Pred. No. 2.8e-15;  
 Matches 198; Conservative 130; Mismatches 294; Indels 233; Gaps 36;  
 DB 4 DKALCGSLTWTQTFHVPSPCAS--PQDLSSGLAVA-----YVLNQIDPSPWFNEAWLQ 55  
 DB 139 DGLNLEDNLEFLQKAPVSTCSSTFPEELSPSHQAKREIRLELQKVASSSSGNFLS 198  
 DB 56 GISEDGPNWKLKVNKLWVLSVYSQDVLAHPVSEHLPDVSLIGFSDPAELGKLL 115  
 DB 199 GSPASP-----MGDILQTPQPMRLKKQLADERSNRDELE--L 235  
 DB 116 QLVLCGAISECK-----KQDHQRIQWTLSESVQHVME--AIQELMTKDTSPSLSPET 167  
 DB 236 ELAENRKLTEKDAQIATMMQORIDRLALLNEKQAASPLEPKELEEL--RDKNESLTMRHL 293  
 DB 168 GNF-----DSQSRYYFLSEAEEDG-----ELQQRCLDLERQMLLSEKQSL 211  
 DB 294 ETUKQODLTKESQMDRKINQJLSE--ENGDLFSKLPREFASHLQQLQDALNELTEHSA 351  
 DB 212 AQENAGLRMRGPEGEGTGLTAKKLL-----LLQSLEQLQAEENFRLESGREDE 263  
 DB 352 TQE---WLEKQAOLERKLSAALQDKCKLEKNEILQGLKLSQLEHLSQLDNPPQKEGV 408  
 DB 264 -----LRCAELERVAEQLQHN-----QAELTS 287  
 DB 409 LGDVLTQLTLEKQAATLAANTQOARVEMLETERGQOEAKLLAERGHFEERQQLSSLI 468  
 DB 288 QEAQALKDMDLROSSERAGOLE-----ATLSCRRLGELRELRLQVQLERNA 339  
 DB 469 TDQSSTSNLSQAKKELEQASQANGARLTQAQVASTS-----ELTLNATIQOQOELA 522  
 DB 340 GHARTQLEDELRRAGLSQALEAQ---BRVQELQG---QROEEAMK---AERKLFECRN 392  
 DB 523 GLKQAQKEKAQL--AOTLQOQEQASGLRHQVQSLSSLKQKQEQQLKEVAEKQATROD 580  
 DB 393 LEKYESVTEKEKRLAERDSL---REANEELCAQLQ-----PRLTQ 433  
 DB 581 HAQOLATAABEREASLRERDAALQLEALEKEKAALKEILQOQQLQVANEARDSAQTSVTQ 640  
 DB 434 ADPSLDSTSPVDNLAA-----EILPAELRETLLRLQLENKRLCROEAADR----- 481

DB 641 AQREKAELSRKVEELQACVETARQEQHEAQVAEILELQRLSE---QQKATEKERVQAEK 697  
 DB 482 ---QELQORHLEDANRARGHLETOHR-----LNOOQ--LSELRAQVEDL-----OKAL 524  
 DB 698 DQLOEQOLQAKESLKVTKGSLSEKRRADALEEQORCISELKAEATRSVLEQHKRERKEL 757  
 DB 525 QE-----QGGKTEDAISILLARKLEEHQKLEHLELQKLEHLELQKREYIELEPPT 571  
 DB 758 EERAGRGKLEARLLQLEGAHQAEVTEVLRLELAEMAAOHTAESECE---QLVKEVAANR 814  
 DB 572 DSTARRIEELQHNQKDDADLRAMEERYRYRVYDKARVMYQTMPEKQRPAA-----622  
 DB 815 DGYEDSQOEEAAYGAMFQQLMTLKEE---CEKARQELQ--EAEKVAGIESHSELQI 867  
 DB 623 ----GAPPELHS-----LRTQLRERDVRIHLEMDF-----EKSRSQREQEKKLLISAWY 670  
 DB 868 SRQONKLAELHANLALQAOVQEKVRAQLADDLSTLQAEKMAATSKAVARL-----919  
 DB 671 GMALOQAGEERAPA 685  
 DB 920 -ETLVKAGEQOETA 933  
 RESULT 8  
 US-09-452-294-1  
 ; Sequence 1, Application US/09452294  
 ; Patent No. 6287790  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lellievre, Sophie  
 ; APPLICANT: Bissell, Mina  
 ; TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED  
 ; TITLE OF INVENTION: THERAPY AND DETECTION OF PROLIFERATIVE AND  
 ; TITLE OF INVENTION: DIFFERENTIATION DISORDERS  
 ; FILE REFERENCE: IB-1454- Sequence Submittal  
 ; Patent No. 6287790  
 ; CURRENT APPLICATION NUMBER: US/09/452,294  
 ; CURRENT FILING DATE: 1999-11-30  
 ; PRIOR APPLICATION NUMBER: 60/110,420  
 ; PRIOR FILING DATE: 1998-11-30  
 ; NUMBER OF SEQ ID NOS: 1  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 2101  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-452-294-1  
 Query Match 8.6%; Score 311.5; DB 4; Length 2101;  
 Best Local Similarity 23.2%; Pred. No. 2.8e-15;  
 Matches 198; Conservative 130; Mismatches 294; Indels 233; Gaps 36;  
 DB 4 DKALCGSLTWTQTFHVPSPCAS--PQDLSSGLAVA-----YVLNQIDPSPWFNEAWLQ 55  
 DB 139 DGLNLEDNLEFLQKAPVSTCSSTFPEELSPSHQAKREIRLELQKVASSSSGNFLS 198  
 DB 56 GISEDGPNWKLKVNKLWVLSVYSQDVLAHPVSEHLPDVSLIGFSDPAELGKLL 115  
 DB 199 GSPASP-----MGDILQTPQPMRLKKQLADERSNRDELE--L 235  
 DB 116 QLVLCGAISECK-----KQDHQRIQWTLSESVQHVME--AIQELMTKDTSPSLSPET 167  
 DB 236 ELAENRKLTEKDAQIATMMQORIDRLALLNEKQAASPLEPKELEEL--RDKNESLTMRHL 293  
 DB 168 GNF-----DSQSRYYFLSEAEEDG-----ELQQRCLDLERQMLLSEKQSL 211  
 DB 294 ETUKQODLTKESQMDRKINQJLSE--ENGDLFSKLPREFASHLQQLQDALNELTEHSA 351  
 DB 212 AQENAGLRMRGPEGEGTGLTAKKLL-----LLQSLEQLQAEENFRLESGREDE 263  
 DB 352 TQE---WLEKQAOLERKLSAALQDKCKLEKNEILQGLKLSQLEHLSQLDNPPQKEGV 408  
 DB 264 -----LRCAELERVAEQLQHN-----QAELTS 287





Patent No. 5906819  
GENERAL INFORMATION:  
APPLICANT: Kaibuchi, Kojo  
APPLICANT: Iwamatsu, Akihiko  
APPLICANT: Nakano, Takeshi  
APPLICANT: Ito, Masaaki  
APPLICANT: Takahashi, No. 5906819uaki  
TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESS: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,576  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-325129  
FILING DATE: 20-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-17150  
FILING DATE: 05-JAN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-131206  
FILING DATE: 26-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16887/843  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-685-576-4

Query Match 8.3%; Score 301; DB 2; Length 1388;  
Best Local Similarity 23.1%; Pred. No. 1e-14;  
Matches 165; Conservative 113; Mismatches 220; Indels 216; Gaps 32;  
QY 151 QELMKTPTSLSPETVGNFDSQSRYYFLSEAEDELEQRCCLDLEROLMLSEKQS 210  
DB 418 ENLLSD-----SPSCRENDIOSRK-----NESEQIOKKLYLLEHLSNEMQAKEE 465  
QY 211 LAQENAGLRMRGPEEGTGLTAKLLLLQSOLEQLENFLESGRDERLRCAELE 270  
DB 466 LEQCKSVNTRLEK-----TAKEL-----EEITLRKSVL-SALR---QLE 502  
QY 271 REVAELQHRN-----QALTSLAQEAQALKDEMLRQ-----SSERAGOLEA 312  
DB 503 REKALLQHKNAEYQKADHEADKRRLENDVNSLKQLEDLKKRNQNSQISTEKVNQQR 562  
QY 313 TL-----TSORRRIGELRRLRRQVRQLEERN----- 338  
DB 563 QLDENALLRSTDAARLRKTQ-AESSKIQQLSENNRDLQDKNCLLETAKLKEFEI 621  
QY 339 -----AGHAEPTRO-----LEDELRRAGSLAQLEAQRQVQE----- 371  
DB 622 NLOSALESERDRTHGSIINDLOGRICGLEEDLKNGLKILLAKVLELEKRLQQRFTDLEK 681

QY 372 -----LQOROEAA-MKAEKWLFECRN-----LEEKYESVTKEKERLIA 409  
DB 682 EKSNEIDMTYOLKYVQOSLEOEAEHAKATKARLADKNKIYESIEEAKSEAKMEKKLL 741  
QY 410 ERDSLREANEEL-----RCAQLQPRGLTQADPSLDPTSTPYDNL--AAEILPAELRETL 461  
DB 742 EERTLKQKVENLLLEAEKCSLL-----DCDLKQSQOKINELLKQKDVNLNEDVRNLT 793  
QY 462 LRLQLENKRLCRQEAADRERQBEL-----QRHLEDANRARGHLETOHRLNOQQLSELRA 515  
DB 794 LKIEQETQKRCILTONDKMQQVNTLKMSEKQLQKEN--NHLMEMKMNLERQN-AELRK 850  
QY 516 QVEDL---OKALOEQGGKTEDAISILLK---PKLEHLQKLHEADLELQKREYIEELP 569  
DB 851 ERDADGQKMELODQ-LEAEQYFSTLYKTQVRELKECEKTKLQKELQCKK---QELQD 906  
QY 570 PTDSSTARRIEELQHNLOKQKADADLAMEERYRYVD-KARVMQTMPEKQRPAGAPPEL 628  
DB 907 ERDSLAA---QLEITLTAKADSEQLARSIAEQSYDLEKEKIMKELEKEMMAR----- 956  
QY 629 HSLRTQLRERDVRIRHLE-----MDFEKSRSQREQEKKLLI 664  
DB 957 H--KQELTEKDATIASLEETNRTLTSDVANLANEKEBELNKNKADVQQLSKLDEE---I 1011  
QY 665 SAWYNGMALQORAGEERAPAHQAQSLAQLAORLATNSRRGPLGRSLASLNLRPDTK 718  
DB 1012 SAAAKAQEFKOLLTERTLTQAVNKLAE-----IMNRKEPVKRGNDTVRRREK 1061  
RESULT 11  
US-08-353-700-1  
Sequence 1, Application US/08353700  
Patent No. 5599919  
GENERAL INFORMATION:  
APPLICANT: YEN, TIMOTHY J.  
APPLICANT: RATTNER, JEROME B.  
TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
TITLE OF INVENTION: AND METHODS OF USE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
STREET: 1601 MARKET STREET, SUITE 720  
CITY: PHILADELPHIA  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,700  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REED, JANET E.  
REGISTRATION NUMBER: 36,252  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:

ORGANISM: HUMAN  
US-08-353-700-1

Query Match 8.2%; Score 297; DB 1; Length 3248;  
Best Local Similarity 21.9%; Pred. No. 6.6e-14;  
Matches 160; Conservative 128; Mismatches 260; Indels 182; Gaps 28;

QY 92 SEEHLPDVLIGFSDPAELGKLLQLVLGCAISCKKODHRIQIMTLEESVQHVMVAIQ 151  
DB 1905 SKLHLOEVQMTKIEACIELEKIVG-----ELKKEN-----SDLSEKLEYFSCDHOE 1951  
QY 152 ELMTKDPDLSPTYGNFDSQSRYYFLSBE-----AEGDELOQRCIDLERQMLLSE 206  
DB 1952 LLQVETSEGLNSDLEHAKSSR-----EDIGNVAKVNSKNERFLDVENELSRIS 2005  
QY 207 EKOSLAQENAGLRMRGPEGEGTGLTAKKLLLLQSOLEQLOEENFLESGREDE--L 264  
DB 2006 EKASIEHE-----ALYLEADLEVYVQTEKLCLEKDNENKQKVI 2042  
QY 265 RCALEREV-----AELQHRNQALTSLAQEAQALKDEMDLQ--OSSERAGOLEATL 314  
DB 2043 VCLBEELSVVTSERNQLRGELDTMSKTTALDQJSEKMKETQELSHQSECLHCIOVAE 2102  
QY 315 TSCRRRLGELRELRRQVQLERAGNAGHAERTQLEDELRRAGSLRAQLEAQRQVQELQ 374  
DB 2103 AEVKEKTELLQTLSSDVSELLKDKTHQEKLOSLKDSQALSCLKENQIAQL----- 2157  
QY 375 QROBEAMKAERKLFECRNLEEKYE--SVTKERLLAER-----DSLRAN 418  
DB 2158 NKEKELLVKESLQARLSSESDYEKLVNSKALEALVEKGFALRLSLSTQBEVHOLRRGI 2217  
QY 419 EELRC-----AQLO-PRGLTOADPSLDPTSTPVDNLAAILPAELRETLRLQLENKR 470  
DB 2218 EKLVRVIEADEKKQLHIAEKLERENDSLKDKVENLERLOLQSEENQELVILDAEN-- 2275  
QY 471 LCROEAADRERQEEQLRHLEDANRARGHLETQHRINQOOLSELRAQVEDLQALQEOGGK 530  
DB 2276 -----SKAEVETLKTQIEEMARSL-----KIFELDLVTLRSEKENLTKQIOKOGQ 2321  
QY 531 TED-----AISILLKRR-----LEHLOKLHEA-----DLELQRR 561  
DB 2322 LSELDKLLSFKSLLEKEQAEIQIKESKTAVEMLOLQNELNEAVALQDQOEIMKAT 2381  
QY 562 EYIELEPPTDSSARR--IEELOHNL-----OKKAD-----LRAMEERYR 602  
QY 603 YVDRKVMQTMPEKQPAAGAPPELHSLRQLRERDVRIRHLEMDFEKRS----- 654  
DB 2440 ELEIAR--TNOEHAALAEANSKGEVETLKKIEGWTOSLGLDVTYIRSEKENLTNE 2496  
QY 555 -QROE-----KLISAWYNGMALQORAGERAPAHQAQSLAQRATNSRRGPLGRLA 709  
DB 2497 LQKQERISELEIINSSFENT--LOEKE-QEKVQMKESKTAMEMLOT-----OLK 2544  
QY 710 SLNLRPTDKH 719  
DB 2545 ELNERVAALH 2554

RESULT 12  
PCT-US95-16216-1  
Sequence 1, Application PC/TUS9516216  
GENERAL INFORMATION:  
APPLICANT: Yen, Timothy J.  
APPLICANT: Rattner, Jerome B.  
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
EXPRESSED KINETOCORE PROTEIN, AND METHODS OF USE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dann, Dorfman, Herrell and Skillman  
STREET: 1601 Market Street Suite 720  
CITY: Philadelphia

STATE: PA  
COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16216  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/353,700  
FILING DATE: 09-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36,252  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US95-16216-1

Query Match 8.2%; Score 297; DB 5; Length 3248;  
Best Local Similarity 21.9%; Pred. No. 6.6e-14;  
Matches 160; Conservative 128; Mismatches 260; Indels 182; Gaps 28;

QY 92 SEEHLPDVLIGFSDPAELGKLLQLVLGCAISCKKODHRIQIMTLEESVQHVMVAIQ 151  
DB 1905 SKLHLOEVQMTKIEACIELEKIVG-----ELKKEN-----SDLSEKLEYFSCDHOE 1951  
QY 152 ELMTKDPDLSPTYGNFDSQSRYYFLSBE-----AEGDELOQRCIDLERQMLLSE 206  
DB 1952 LLQVETSEGLNSDLEHAKSSR-----EDIGNVAKVNSKNERFLDVENELSRIS 2005  
QY 207 EKOSLAQENAGLRMRGPEGEGTGLTAKKLLLLQSOLEQLOEENFLESGREDE--L 264  
DB 2006 EKASIEHE-----ALYLEADLEVYVQTEKLCLEKDNENKQKVI 2042  
QY 265 RCALEREV-----AELQHRNQALTSLAQEAQALKDEMDLQ--OSSERAGOLEATL 314  
DB 2043 VCLBEELSVVTSERNQLRGELDTMSKTTALDQJSEKMKETQELSHQSECLHCIOVAE 2102  
QY 315 TSCRRRLGELRELRRQVQLERAGNAGHAERTQLEDELRRAGSLRAQLEAQRQVQELQ 374  
DB 2103 AEVKEKTELLQTLSSDVSELLKDKTHQEKLOSLKDSQALSCLKENQIAQL----- 2157  
QY 375 QROBEAMKAERKLFECRNLEEKYE--SVTKERLLAER-----DSLRAN 418  
DB 2158 NKEKELLVKESLQARLSSESDYEKLVNSKALEALVEKGFALRLSLSTQBEVHOLRRGI 2217  
QY 419 EELRC-----AQLO-PRGLTOADPSLDPTSTPVDNLAAILPAELRETLRLQLENKR 470  
DB 2218 EKLVRVIEADEKKQLHIAEKLERENDSLKDKVENLERLOLQSEENQELVILDAEN-- 2275  
QY 471 LCROEAADRERQEEQLRHLEDANRARGHLETQHRINQOOLSELRAQVEDLQALQEOGGK 530  
DB 2276 -----SKAEVETLKTQIEEMARSL-----KIFELDLVTLRSEKENLTKQIOKOGQ 2321  
QY 531 TED-----AISILLKRR-----LEHLOKLHEA-----DLELQRR 561  
DB 2322 LSELDKLLSFKSLLEKEQAEIQIKESKTAVEMLOLQNELNEAVALQDQOEIMKAT 2381  
QY 562 EYIELEPPTDSSARR--IEELOHNL-----OKKAD-----LRAMEERYR 602

Db 2382 E--QSLDPPIEEHQLRNSTKRLARLEADEKKQLCVLQOLKESEHDLKGRVNLER 2439  
 QY 603 YVDKARVMOTMPKORPAAGAPPELHSLRTQLRERDVIRIRHLEMDFKRSR 654  
 Db 2440 ELIAR--TNQHAALAEANSKGEVETLKAKTEGWTQSLRGLDVTIRSEKENLTNE 2496  
 QY 655 -OREQEE---KLLISAWYNGMALQORAGEERAPAHQSFQAQRLATNRRRGPGLRLA 709  
 Db 2497 LQKEORISELEINSSFENI---LQKE-QEKVQMKESSTANEMLOT-----OLK 2544  
 QY 710 SLNLRPTDKH 719  
 Db 2545 ELNERVAALH 2554

## RESULT 13

US-08-328-254-6  
 ; Sequence 6, Application US/08328254  
 ; Patent No. 5710022  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zhu, Xueliang  
 ; APPLICANT: Lee, Wen-Hwa  
 ; TITLE OF INVENTION: A No. 5710022a Nuclear Mitotic Phosphoprotein  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Campbell and Flores  
 ; STREET: 4370 La Jolla Village Drive, Suite 700  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 92122

COMPUTER TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/328,254  
 FILING DATE: 24-OCT-1994  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/141,239  
 FILING DATE: 22-OCT-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Campbell, Cathryn A.  
 REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-CJ 1191  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 535-9001  
 TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2482 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-328-254-6

Query Match 8.1%; Score 296; DB 1; Length 2482;  
 Best Local Similarity 21.9%; Pred. No. 5.5e-14;  
 Matches 160; Conservative 128; Mismatches 260; Indels 182; Gaps 28;

QY 92 SEHLPDVLGIEFSDPRLGKLLQLVLGCAISCEKKODHQIRIMTLEESVQHVVMIAIQ 151  
 Db 1177 SKLHLQEQVQLMTKIEACIEKIVG-----ELKKN-----SDLSEKLEFFSCDHOE 1223  
 QY 152 ELMTKDTPLSLPETYGNFDSQSRYYFLSEE-----AEEGDELQORCLDLERQLMLLSE 206  
 Db 1224 LQRVETSEGLSNDLENHAKDSR-----EDIGDNVAKVNSWKERFLDVNELSRIRS 1277  
 QY 207 EKQSLAQENAGLRMRGPEGETPGLTAKKLLQLQSLQEQLEQENFRLESGREDE--L 264

Db 1278 EKASIEHE-----ALYLEADLEVVTQTEKLCLEKONENKQKVI 1314  
 QY 265 RQAELEREV-----AELQHRNQALTSLAQEAQALQDEMDLR-QSSERAGOLEATL 314  
 Db 1315 VCLEEELSVTSEGNOLRGELDTMSKTTALDQLSKMKETQELSEHOSCHLCIQVAE 1374  
 QY 315 TSCRRRLGELRELRRVROLEERNAGHAERTROLEDRLRAGSLRAQLEAQRROVQLOQ 374  
 Db 1375 AEVKEITELLTSSDYSELLKDKTHLQEKLOSLQSLQSLQSLQSLQSLQSLQSLQSLQ 1429  
 QY 375 QROEEAMKAFKWLFEGRNLEEKYE--SVTKERKERLIAER-----DSUREAN 418  
 Db 1430 NKEKELLVKESESLQARLSSEDEYKLVNSKALEAALVEKGEFALRLSSSTQEEVHQLRRI 1489  
 QY 419 EELRC-----AQIQ-PRGLTOADPSLDPTSTPDVNLAAEILPAELRETLRLLOLENKR 470  
 Db 1490 EKLVRVREADEKKQLHIAEKLKERENDSLKDKVENLERELQMSSENLQELVILDAEN-- 1547  
 QY 471 LCRQEAADRRERQELQRLHLDANRARRHGLETOHRLNQOQLSELRAQVEDLQALQEQGK 530  
 Db 1548 -----SKAETVLTQIEMARSL-----KVFELDLVTLRSEKENLTQIQEKQOQ 1593  
 QY 531 TED-----AISILLRKK-----LEEHLOKLHEA-----DLEQKR 561  
 Db 1594 LSELDKLLSFKSLLEKEQAEIQKEESTAVEMLQNLKELNEAVALCGDOEIMKAT 1653  
 QY 562 EYIEEPPPTDSSTARR--TEELOHNL-----OKKQAD-----LRAMEERYR 602  
 Db 1654 E--QSLDPPIEEHQLRNSTKRLARLEADEKKQLCVLQOLKESEHDLKGRVNLER 1711  
 QY 603 YVDKARVMOTMPKORPAAGAPPELHSLRTQLRERDVIRIRHLEMDFKRSR----- 654  
 Db 1712 ELEIAR--TNQHAALAEANSKGEVETLKAKTEGWTQSLRGLDVTIRSEKENLTNE 1768  
 QY 655 -OREQEE---KLLISAWYNGMALQORAGEERAPAHQSFQAQRLATNRRRGPGLRLA 709  
 Db 1769 LQKEORISELEINSSFENI---LQKE-QEKVQMKESSTANEMLOT-----OLK 1816  
 QY 710 SLNLRPTDKH 719  
 Db 1817 ELNERVAALH 1826

## RESULT 14

US-08-685-576-1  
 ; Sequence 1, Application US/08685576  
 ; Patent No. 5906819  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kalebuchi, Koza  
 ; APPLICANT: Iwamatsu, Akihiro  
 ; APPLICANT: Nakano, Takeshi  
 ; APPLICANT: Ito, Masaaki  
 ; APPLICANT: Takahashi, No. 5906819uaki  
 ; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
 ; NUMBER OF SEQUENCES: 16  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/685,576  
 ; FILING DATE: 24-JUL-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 7-325129

FILING DATE: 20-NOV-1995  
PRIOR APPLICATION DATA: JP 8-17150  
FILING DATE: 05-JAN-1996  
PRIOR APPLICATION DATA: JP 8-131206  
FILING DATE: 26-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16887/843  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-685-576-1

Query Match 8.1%; Score 295; DB 2; Length 1388;  
Best Local Similarity 22.5%; Pred. No. 3e-14;  
Matches 155; Conservative 126; Mismatches 241; Indels 168; Gaps 30;

Qy 151 QELMTKPTDPSLSPETYNFDSQRRYYFLSPEAEGBDELQORCLDLERQLMLLSEKQS 210  
Db 418 ENLLSD-----SPCKENDSIQSK-----NESEIQKLLTLEHSTEIOAKEE 465

Qy 211 LAQENAGLRERGRPEGTGLTAKKLLQLSQLEIQEE-----NFRLESGRED 261  
Db 466 LEQCKSVNTRLEKVAKELEBEITLTK-NVESTLRQLEREKALLQHNKAEVQKADHEA 523

Qy 262 EELRCAE-----LREVAELQHRNQALTSLOEAQALKDEWDE-----LRQSSERAGOLEA 312  
Db 524 DKRNLENVNSLKDQLEDLKKRNQNSISTEKNVQLQRLDETALLRTESDTAARLK 583

Qy 313 TLTSRRRLGEL-----RELK-----ROVROLEERNAGHAE----- 343  
Db 584 TQAESSKIQIOLESNNRDLQKNCLETAKLKEKEFINLOSLESERRDRTHGSEIND 643

Qy 344 ---RTROLEDELRLAGSLRAQLEAQRQVQE-----LQGORQ 377  
Db 644 LQGRISGLEEDVKNKILLAKVELEKRLQSRFTDLREKKNMEIDMTYQLKVIOQSLEQ 703

Qy 378 EEA-MKAEKWLFECHN-----LEKYESVTYKEKRLLEABRSLREANSEL-----RCAQ 425  
Db 704 EETEHKATKARLADKNKYIESEAKSEAMKEMKKLSEERTLKOKVENLLEAEKRCGI 763

Qy 426 LQPRGLTQADPSLQPTSTPVDNL--AAEILPAELRETLRLQLENKRLCROEAADRERQE 483  
Db 764 L-----DCDLKQSOQKINELLKQKQVLDVNEVNLTKIEQTKKRLQNDLQKMQTQ 815

Qy 484 EL-----QRHLEDANRAHGLETOHRLNQOOLSELRAQVEDL---QKALQEQGKTEDA 534  
Db 816 QVNTLKMSEKOLKQEN--NHLLEKMSLEKQK-AELRKERQADQOMKELQDQ-LEASQY 871

Qy 535 ISILLK---RLLEHLQKLHLEADLELQKRYIELEPPTSSSTARRIELOHNLQKDDA 591  
Db 872 FSTLYKTQVRELKECECEKTKLCELQOQK---QELQDERDSLAA---OLEITLTKADS 924

Qy 592 DLRAEMEERYRYVD-KARVMOTMEPKORPAAGAPPELHSLRTQLRERDVIRHLE----- 646  
Db 925 EQLARSIAEBOYSLEKIKMKELTEIKEMMAR-----H--KQELTEKDATIASLEETNR 976

Qy 647 ---MDFEKRSQROEEKLLISANYMGMALQORAGRAPAHA-----QSFIAQOQLA 697  
Db 977 TLTSVDANLANKEELNKNKAEQOL-----SRKDEISAIAKQAFEXQLLTERTLK 1031

Qy 698 TN-----SRGPGRLGRASLNLRPDK 718  
Db 998 TN-----SRGPGRLGRASLNLRPDK 718

Db 1032 TQAVNKLAEIMNRKEPVKRGNDTDVRRKEK 1061

RESULT 15  
US-08-533-306A-4  
Sequence 4, Application US/08533306A  
Patent No. 5837457  
GENERAL INFORMATION:  
APPLICANT: Liu, Pu  
APPLICANT: Collins, Francis S.  
APPLICANT: Siciliano, Michael J.  
APPLICANT: Claxton, David  
TITLE OF INVENTION: Markers for Detection of Chromosome 16  
TITLE OF INVENTION: Rearrangements  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
STREET: P.O. Box 828  
CITY: Bloomfield Hills  
STATE: MI  
COUNTRY: USA  
ZIP: 48303

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/533.306A  
FILING DATE: September 25, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Deann F.  
REGISTRATION NUMBER: 36683  
REFERENCE/DOCKET NUMBER: 2115-00869COB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810) 641-1600  
TELEFAX: (810) 641-0270  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 885 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-533-306A-4

Query Match 7.8%; Score 284.5; DB 2; Length 885;  
Best Local Similarity 20.8%; Pred. No. 1e-13;  
Matches 160; Conservative 127; Mismatches 257; Indels 225; Gaps 33;

Qy 119 LGCAISCEKKQDHQIRIMTLEESVQHVYVWEAIQELMTKDTDPDSLSPETYNFDSQRRYY 178  
Db 121 MGC---LEFDEERACQEDALAQAFEAARRRTREFDRDRSHREEMAEKANLDKNKQ--- 174

Qy 179 FLSEAEEDG-----ELQORCLDLERQLMLLSEKQS 211  
Db 175 --TLEKENADLAGELRVLGQAKQVEHKKKLEAQVQLQSKCSDGERARAEKNDKVHKL 232

Qy 212 AQENAGLRERGRPEGTGTLTAKKLLQLSQLEQENFRLESGREDERLRCAELER 271  
Db 233 QNEVESVTGMLNEAEGRAIK--LAKDVASLSQLQDTQE-----LLQE 273

Qy 272 EVAELOHNRQALTSLOEAQALKDEMD---ELROSSER-AQGLEATLTSRRRLGEL--- 324  
Db 274 ETROKLVNVTKLQLEERNSLOQDDEMEAKONLHISTLNQLSDSKKLQDFAST 333

Qy 325 -----RELRCVROLEERNAGH---AETR-QLEDELRRAGSLRAQLEAQRQV 369  
Db 334 VEALLEGKKRQKEIENLTQOYEKAAYDKLETKRKLQOEL---DDLVDLDNQRLV 390

Qy 370 QELQORQOEAEKAEKWLFECHN-----ESVTKKE-----RLAE---R 411  
Db 391 SNLEKKQR---KFDQLLAEEKNISSKYADDERDRAEFAAREKETKALSARALEALEAK 446

Wed Mar 5 08:37:50 2003

```

QY 412 DSIREANEELCAQLQ-----PRGLTQADPSLPTSTPVDNLAEEILPA 455
Db 447 EELERTNKMLK-AEMEDLVSSKDDVGNVHELEKSKRALETONEEKTKOLELEDELOAS 505
QY 456 ELRETLRLLOENKRLCRO-----EAADREROE---ELQRHLED-----AN 493
Db 506 E--DAKLRLEVNVQALKGOFERDLOARDQONEEKRRLQRLQRLHEVETELEDERNERAA 563
QY 494 RARHGLETOHR-LNOQ-----OLSELRAQVEDLOKALQEQGKGTEDAISILL 539
Db 564 AAKKLEGLDKDLELOADSARKGREAIRKLRKLOAQMKDFORELEDAASRDEIFATAK 623
QY 540 K-----RKLEEHLOKLHE-----ADLELQKREYIEELEPPPTDSTA----- 576
Db 624 ENKKAKSLEADLMQOEDLAAERARKQADLE---KEELAEELASSLSGRNALQDEKRR 680
QY 577 --RRIEELQHNQKQDADLRAMEERYRYRYVDKARMVMOTMEPKORPA---AGAPPELH-- 629
Db 681 LEARIALEEELEEEGNGMEAMSDRVKATQQAEOQLSNELATERSTAQKNESARQOLERO 740
QY 630 --SLRTOLRR-----DVRIRHLEMDFEK-----SRSOREOEKL-- 662
Db 741 NKELRSKLHEMEGAVKSKFKSTIAALEAKIAQLEEQVEQAREKQAKTKSLKQKDKLKE 800
QY 663 -LISAWYNGMALQORAGERERAPAHQSFQAO-----ORLATNSRR 702
Db 801 ILLOVEDERKMAEQYKEAQEKGNARVKQLKRLQLEEAEEESQRINANPRK 849

```

Search completed: March 4, 2003, 15:06:22  
Job time : 38.9261 secs

| Result No. | Score | Query Match | Length | DB | ID                  | Description           |
|------------|-------|-------------|--------|----|---------------------|-----------------------|
| 1          | 315   | 8.7         | 2310   | 9  | US-09-991-496-120   | Sequence 120, App     |
| 2          | 315   | 8.7         | 2310   | 9  | US-09-820-843A-114  | Sequence 114, App     |
| 3          | 315   | 8.7         | 2310   | 10 | US-09-874-923-120   | Sequence 120, App     |
| 4          | 315   | 8.7         | 2354   | 9  | US-09-820-843A-113  | Sequence 113, App     |
| 5          | 299   | 8.2         | 1945   | 9  | US-09-927-597-2     | Sequence 2, Appli     |
| 6          | 299   | 8.2         | 1979   | 9  | US-09-927-597-4     | Sequence 4, Appli     |
| 7          | 290   | 8.0         | 1958   | 12 | US-10-028-946-4     | Sequence 4, Appli     |
| 8          | 290   | 8.0         | 2053   | 9  | US-10-017-216-2     | Sequence 2, Appli     |
| 9          | 290   | 8.0         | 2054   | 12 | US-10-028-946-2     | Sequence 2, Appli     |
| 10         | 290   | 8.0         | 2055   | 9  | US-10-017-216-4     | Sequence 4, Appli     |
| 11         | 284.5 | 7.8         | 868    | 9  | US-09-884-001-19    | Sequence 19, Appli    |
| 12         | 282   | 7.8         | 1641   | 9  | US-10-017-216-5     | Sequence 5, Appli     |
| 13         | 277   | 7.6         | 1162   | 10 | US-09-815-242-11828 | Sequence 11828, Appli |
| 14         | 276   | 7.6         | 1597   | 9  | US-10-017-216-6     | Sequence 6, Appli     |
| 15         | 270.5 | 7.4         | 879    | 9  | US-10-108-605-217   | Sequence 217, App     |
| 16         | 269.5 | 7.4         | 2139   | 9  | US-10-023-219-4     | Sequence 4, Appli     |
| 17         | 269.5 | 7.4         | 2139   | 10 | US-09-727-384-6     | Sequence 6, Appli     |
| 18         | 269   | 7.4         | 2125   | 10 | US-09-919-172-29    | Sequence 29, App      |
| 19         | 266   | 7.3         | 900    | 12 | US-10-071-751-21    | Sequence 21, App      |

306 RAGOLEATLTSRRRLGELRELRRQ-----VROLEERNAGHAERTQLEDELRR-----354  
Db 868 TATQORAELEARVARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLARLAARD 927  
QY 355 -----AGSLRAQLEAQRQVQELQO-----ROEAMKAQKWLFEKNLE 394  
Db 928 EARGQOLAAAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQ 980  
QY 395 EKVESYVTEK-----ERLLAERDSLRE-----ANEELRCAQLQPRGLTQADPSLDTSTP 444  
Db 981 QRLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQOR-----LDTATQO 1028  
QY 445 VDNLAEEI--LPALRETLRLQLENKRLCRQ-----AADRE--RO-----EELQRH 488  
Db 1029 RALEAQLARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQOR 1088  
QY 489 LEDANRARGLET-----OHRNLN--OOQLSELRAQV-----EDQKALQ-- 525  
Db 1089 LDTATQORAELEARVARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQOR 1148  
QY 526 -EGGKTEDAISIL-----LKRKLEEHLOKLHLEADLELQORREYIEEL 567  
Db 1149 TQORAELEARVARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQOR 1208  
QY 568 EPTDSTARRIEELQHNQKQDAD-----LRAMEERYRYVYDKARVMQTMPEK-QR 619  
Db 1209 QORLDTATQOR--AELEAQLARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLAR 1267  
QY 620 PAAGAPPELHSLTQLRER---DVRIRHLEMDPEKRSQ-----REOEKLLISAWYNGM 671  
Db 1268 LAANAEELQORLDTATQORAELEARVARLAADRDDEARQOLAAAEELQORL-----D 1319  
QY 672 MALQORAGEERAPAH--AQSFLLAQORLATNS 700  
Db 1320 TATQORAELEARVARLAADRDDEARQOLAAAEELQORL-----D 1319

RESULT 2  
US-09-820-843A-114  
; Sequence 114, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN  
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES  
; FILE REFERENCE: 063915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 114  
; LENGTH: 2310  
; TYPE: PRT  
; ORGANISM: L. major  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: AC005893\_12 L6202.3  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: gi|6899664  
US-09-820-843A-114

Query Match 8.7%; Score 315; DB 9; Length 2310;  
Best Local Similarity 25.5%; Pred. No. 5e-10;  
Matches 176; Conservative 96; Mismatches 205; Indels 214; Gaps 32;

QY 184 ABEDELOQRC-----LDLEROIMLS-----BEKQSLA-----QENAGL 218  
Db 700 AANAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQORLDTATQORAELE 759  
QY 219 RERMGPEGCTPGTAKKLLLSOLEQOEFNFRLESGREDELRCAELEREVAEL-Q 277  
Db 760 EARVARLAADGDDEARQO-----LAANAEELQO-----RLDTATQOR-----AELEARVARLA 807

QY 278 HRNOALTSIAQBAQALKDBMD-----ELQRSSE 305  
Db 808 DDEARQOLAAAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQORL 867  
QY 306 RAGOLEATLTSRRRLGELRELRRQ-----VROLEERNAGHAERTQLEDELRR-----354  
Db 868 TATQORAELEARVARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLARLAADRD 927  
QY 355 -----AGSLRAQLEAQRQVQELQO-----ROEAMKAQKWLFEKNLE 394  
Db 928 EARGQOLAAAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQOR-----LQ 980  
QY 395 EKVESYVTEK-----ERLLAERDSLRE-----ANEELRCAQLQPRGLTQADPSLDTSTP 444  
Db 981 QRLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQOR-----LDTATQO 1028  
QY 445 VDNLAEEI--LPALRETLRLQLENKRLCRQ-----AADRE--RO-----EELQRH 488  
Db 1029 RALEAQLARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQOR 1088  
QY 489 LEDANRARGLET-----OHRNLN--OOQLSELRAQV-----EDQKALQ-- 525  
Db 1089 LDTATQORAELEARVARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQORLDTA 1148  
QY 526 -EGGKTEDAISIL-----LKRKLEEHLOKLHLEADLELQORREYIEEL 567  
Db 1149 TQORAELEARVARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLARLAADRDDEARQOLAAAEELQOR 1208  
QY 568 EPTDSTARRIEELQHNQKQDAD-----LRAMEERYRYVYDKARVMQTMPEK-QR 619  
Db 1209 QORLDTATQOR--AELEAQLARLAADRDDEARQOLAAAEELQORLDTATQORAELEAQLAR 1267  
QY 620 PAAGAPPELHSLTQLRER---DVRIRHLEMDPEKRSQ-----REOEKLLISAWYNGM 671  
Db 1268 LAANAEELQORLDTATQORAELEARVARLAADRDDEARQOLAAAEELQORL-----D 1319  
QY 672 MALQORAGEERAPAH--AQSFLLAQORLATNS 700  
Db 1320 TATQORAELEARVARLAADRDDEARQOLAAAEELQORL-----D 1319

RESULT 3  
US-09-874-923-120  
; Sequence 120, Application US/09874923  
; Patent No. US20020081320A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Webb, John R.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Coler, Rhea  
; APPLICANT: Probst, Peter  
; APPLICANT: Brannon, Mark  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
; FILE REFERENCE: 210121.420C8  
; CURRENT APPLICATION NUMBER: US/09/874,923  
; CURRENT FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 122  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 120  
; LENGTH: 2310  
; TYPE: PRT  
; ORGANISM: Leishmania major and chagasi  
US-09-874-923-120

Query Match 8.7%; Score 315; DB 10; Length 2310;  
Best Local Similarity 25.5%; Pred. No. 5e-10;  
Matches 176; Conservative 96; Mismatches 205; Indels 214; Gaps 32;

| Query Match           | 8.7%                  | Score 315;            | DB 9;                   | Length 2354;                 |
|-----------------------|-----------------------|-----------------------|-------------------------|------------------------------|
| Best Local Similarity | 26.1%                 | Pred. No. 5.1e-10;    |                         |                              |
| Matches               | 184;                  | Conservative          | 99;                     | Mismatches 251;              |
|                       |                       |                       |                         | Indels 170;                  |
|                       |                       |                       |                         | Gaps 34;                     |
| QY 139                | EESVQHVMEAIQELMYKDT   | PDLS                  | SPETYGNFDSQSRYYFLS      | BEAEE---GDELOQR 193          |
| DB 782                | DEARQOLAAANEELQORLDT  | ---                   | ATQRAELEQAVARLAADR      | DEARQOLAAANEELQOR 837        |
| QY 194                | C-----LDLEQLMLIS      | ---                   | EKOSLAGENAGLRMG         | -----RPEGSTPGLTAKK 237       |
| DB 838                | LDATQORAELEQAVARLAADR | DEARQOLAAANEELQORLDT  | ATQORAELEQAVARLAAN      | 897                          |
| QY 238                | LLLLQSLE-----         | QLQENFLESGREDELR      | CAELREVEAELQHRNQA       | LYSLAQEA 290                 |
| DB 898                | AEELQRLDTATQORAELEA   | VARLAADR              | DEAR---QOLAAANEELQORLDT | ATQORAE 954                  |
| QY 291                | QA-----LKDNDLELRQS    | SERAGOLEATLT          | SCRRRLGELR              | ELAROVQRLERNAGHAER 344       |
| DB 955                | EAQLARLAADR           | DEARQOLAAANEELQORLDT  | ATQORAELEAQLARLAADR     | DEARQOLAAAN 1014             |
| QY 345                | TQLEDELRRAGSLARQALEA  | Q                     | RRQVQELQORQEEAMKA       | KWFLFECNLBEKYESVTKE 403      |
| DB 1015               | AEELQORLDTATQORAELEA  | QAVARLAADR            | DEARQOLAAANE-----       | LQORLDTATQO 1067             |
| QY 404                | K-----ERLLAERDSURE    | ---                   | ANEEELRCAQLQPRGLT       | QADPSLDTSTPYVDNLAA--- 450    |
| DB 1068               | RAELEARVARLAADR       | DEARQOLAAANE-----     | ELOQR                   | LDTATQORAELEQAVARLAADGD 1122 |
| QY 451                | --EILPAELRETLLQL      | ENKRLCRQE             | -----AADRE--RQ-----     | EELORHLEDAN 493              |
| DB 1123               | EARQOLAAANEELQORLDT   | ATQORAELEAVARLAADR    | DEARQOLAAANEELQORLDT    | AT 1182                      |
| QY 494                | RARHGLET-----         | ---                   | QHRLN--QOOLSELRAQV      | -----EDLOKAL 524             |
| DB 1183               | QORAELEAQLARLAADR     | DEARQOLAAANEELQORLDT  | ATQORAELEQAVARLAADG     | DEAR 1242                    |
| QY 525                | QEOGGKTEDATSI         | LLKKRLBHLQKLHEAD      | LELQR-----KREYI-----    | EELEPPDPS 573                |
| DB 1243               | QOLAAANE-----         | LQORLDT               | ATQORAELEAQLARLAADR     | DEARQOLAAANEELQORLDT 1297    |
| QY 574                | STARRIEQLHNLOK        | KDADLRAMEERYRYDK      | -----                   | ARMVQTMPEKQRP 621            |
| DB 1298               | ATQORAE-----          | LEAQAVARLAANEELQORLDT | ATQORAELEAVARLAADR      | DEARQOL 1352                 |
| QY 622                | AGAPPELHSLR           | QTLQR---              | DVRIRHLEMDPEKRSQ        | -----REQEEKLISAWYNGMA 673    |
| DB 1353               | ANAEELQORLDT          | ATQORAELEAVARLAADR    | DEARQOLAAANEELQORL      | -----DTA 1404                |
| QY 674                | LOQRAGEE-----         | RAPAHAQSFLAQR         | LL--ATNSRRGPGLRLASL     | 711                          |
| DB 1405               | TQORAELEQAVARLAANEEL  | ---                   | QORLDT                  | ATQORAELEAVARL 1446          |

```

RESULT 5
US-09-927-597-2
; Sequence 2, Application US/09927597
; Publication No. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Fady
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CYTOP018
; CURRENT APPLICATION NUMBER: US/09/927,597
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1945
; TYPE: PRT

```



ORGANISM: Human  
US-09-927-597-4

Query Match 8.2%; Score 299; DB 9; Length 1945;  
Best Local Similarity 24.5%; Pred. No. 3.3e-09;  
Matches 179; Conservative 116; Mismatches 277; Indels 158; Gaps 34;

```
QY 68 KVSNNKWLRLSVESQVLAHPVSE-----EHLPDV-SLIGFSDPAELGKLLQVLGC 121
DB 1084 QIAELKMLAKKEEELQALALRDLDEIAKNNALAKKIRELEGHISD---LOEDLSERAA 1140
QY 122 AISCCKQ----DHQIRMT-LEESVQHVVMIAIQLMTKTPD-----SLSPETYGNF 170
DB 1141 RNKAEKQKRDLCGELEALKTELEDTLDSATQ--QELRAKREQEVTVLKALDEETRSH- 1197
QY 171 DQSRRYFSLSEAEGBDELOQRCIDLERQLMLLSEKQSLAQENAGLRMRGRPEGET 230
DB 1198 --EAOVQEMRQKHAQAVEELTQEQFRAKANLKNKOTLEKENADLAGEL-RVLGQAK 1254
QY 231 PGLTAKKLLLSQLEQOEENFRLESGREDERLRCALEREVAELQHRNQALTSIAQEA 290
DB 1255 QEVHKK-KKLEAQVQELQSK-----CSDGERAR-AELNDKVHKLQNEVESVTGMLNEA 1306
QY 291 Q--ALKDMDLRQSSERAGQLEATLTSRRRLG---ELRELRQVROLEERNAGHAERT 345
DB 1307 ECKAIKAKDVASLSQLODTQELLQEETROKLVNSTKLRLQLEERNSLQDQDDEMEAK 1366
QY 346 RQLEDELRAQSIRQAQLEAQRQVQEQLOQROEEAKAEKWL--ECRNL-----BEK--- 396
DB 1367 QNLE---RHISTINTQLSDSKKKLQDFAS--TVEALEGKKRFQKEIENLTQOYEEKAAA 1421
QY 397 YESVTKEKRLLAERDSLREANEELRCAQLQPRGLTQADPSLDPTSTPVDNLAAE----- 451
DB 1422 YKLEKTKRLQEQELDDLVVDLDNR-----QLVSNLEKKQKQKFDOLLAEEKNIS 1471
QY 452 -----ILPAELRETLRLQLENKRLCQEAADR--ERQELQR-----IELEPPPTD 572
DB 1582 QFERDLQARDEQNEKRRRLQRLQHEYTELEDERQORALAAAKKKLEGDLDKLELQAD 1641
QY 573 SSTARRIEELQHNQKQDADLRAMEERYRYRYVDKARM---VMQTMPEKQRPAGAPPEL 628
DB 1642 SAIKGR-EAIAKQLRK-----LQAKMDFQRELEDAARSDEIFATAKENEKKAASLEADL 1696
QY 629 HSLRTOL-----REDVRIRHLEMDFKRSORE--QEEKLLISAWYNGMALQORAG 679
DB 1697 MLQEDLAAARAKQADLEKELABELASSLSGRNALODEKRRLEARI---AQLEEELE 1753
QY 680 EERAPAHQAS 689
DB 1754 BEQGNMEAMS 1763
```

## RESULT 6

US-09-927-597-4  
Sequence 4, Application US/09927597  
Publication No. US20030032018A1  
GENERAL INFORMATION:  
APPLICANT: Malik, Fady  
APPLICANT: Beraud, Christophe  
APPLICANT: Freedman, Richard  
APPLICANT: Craven, Andrew  
APPLICANT: Sakowicz, Roman  
TITLE OF INVENTION: Human smooth muscle myosin heavy chain  
FILE REFERENCE: CYTOF018

## RESULT 7

US-10-028-946-4  
Sequence 4, Application US/10028946  
Patent No. US20020123622A1  
GENERAL INFORMATION:  
APPLICANT: Yu, Xuanchuan

CURRENT APPLICATION NUMBER: US/09/927,597  
CURRENT FILING DATE: 2001-08-10  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 1979  
TYPE: PRT  
ORGANISM: Human  
US-09-927-597-4

Query Match 8.2%; Score 299; DB 9; Length 1979;

Best Local Similarity 24.5%; Pred. No. 3.3e-09;  
Matches 179; Conservative 116; Mismatches 277; Indels 158; Gaps 34;

```
QY 68 KVSNNKWLRLSVESQVLAHPVSE-----EHLPDV-SLIGFSDPAELGKLLQVLGC 121
DB 1084 QIAELKMLAKKEEELQALALRDLDEIAKNNALAKKIRELEGHISD---LOEDLSERAA 1140
QY 122 AISCCKQ----DHQIRMT-LEESVQHVVMIAIQLMTKTPD-----SLSPETYGNF 170
DB 1141 RNKAEKQKRDLCGELEALKTELEDTLDSATQ--QELRAKREQEVTVLKALDEETRSH- 1197
QY 171 DQSRRYFSLSEAEGBDELOQRCIDLERQLMLLSEKQSLAQENAGLRMRGRPEGET 230
DB 1198 --EAOVQEMRQKHAQAVEELTQEQFRAKANLKNKOTLEKENADLAGEL-RVLGQAK 1254
QY 231 PGLTAKKLLLSQLEQOEENFRLESGREDERLRCALEREVAELQHRNQALTSIAQEA 290
DB 1255 QEVHKK-KKLEAQVQELQSK-----CSDGERAR-AELNDKVHKLQNEVESVTGMLNEA 1306
QY 291 Q--ALKDMDLRQSSERAGQLEATLTSRRRLG---ELRELRQVROLEERNAGHAERT 345
DB 1307 ECKAIKAKDVASLSQLODTQELLQEETROKLVNSTKLRLQLEERNSLQDQDDEMEAK 1366
QY 346 RQLEDELRAQSIRQAQLEAQRQVQEQLOQROEEAKAEKWL--ECRNL-----BEK--- 396
DB 1367 QNLE---RHISTINTQLSDSKKKLQDFAS--TVEALEGKKRFQKEIENLTQOYEEKAAA 1421
QY 397 YESVTKEKRLLAERDSLREANEELRCAQLQPRGLTQADPSLDPTSTPVDNLAAE----- 451
DB 1422 YKLEKTKRLQEQELDDLVVDLDNR-----QLVSNLEKKQKQKFDOLLAEEKNIS 1471
QY 452 -----ILPAELRETLRLQLENKRLCQEAADR--ERQELQR-----IELEPPPTD 572
DB 1582 QFERDLQARDEQNEKRRRLQRLQHEYTELEDERQORALAAAKKKLEGDLDKLELQAD 1641
QY 573 SSTARRIEELQHNQKQDADLRAMEERYRYRYVDKARM---VMQTMPEKQRPAGAPPEL 628
DB 1642 SAIKGR-EAIAKQLRK-----LQAKMDFQRELEDAARSDEIFATAKENEKKAASLEADL 1696
QY 629 HSLRTOL-----REDVRIRHLEMDFKRSORE--QEEKLLISAWYNGMALQORAG 679
DB 1697 MLQEDLAAARAKQADLEKELABELASSLSGRNALODEKRRLEARI---AQLEEELE 1753
QY 680 EERAPAHQAS 689
DB 1754 BEQGNMEAMS 1763
```

```
APPLICANT: Miranda, Maricar
APPLICANT: Fridde, Carl Johan
TITLE OF INVENTION: No. US20020123622A1 Human Kinases and Polynucleotides Encoding
FILE REFERENCE: LEX-0289-USA
CURRENT APPLICATION NUMBER: US/10/028,946
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: US 60/258,335
PRIOR FILING DATE: 2000-12-27
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1958
TYPE: PRT
ORGANISM: homo sapiens
US-10-028-946-4

Query Match 8.0%; Score 290; DB 12; Length 1958;
Best Local Similarity 19.8%; Pred. No. 1.1e-08;
Matches 162; Conservative 137; Mismatches 258; Indels 260; Gaps 28;

QY 92 SEEHLPDV-----SLIGESDPAELGKLLQLVL-----GCAISCEKKQ 129
Db 410 SGEELPFVGSYKALGILGRSESVGLDSPAKTSMKKLLIKSKELQDSQDKCKME 469
QY 130 DHQIRIMTLEESVQHVMEAIQELMTKDPDSLSPETGYNFDQSRRYYFLSEAEGBE 189
Db 470 QEMTRLHRRVSEVAVLSQKEVELKASQTSRSLLEQDLATYITE-----513
QY 190 LQORCLDLERQL-----MLLSEEQSLAQENAGLRNMRGPEGETGLTAKKLLQLSQ 245
Db 514 ----CSSLRSLQARMEVSQEDDKALQLLDHIREQ-----SRKLOEIKQE 556
QY 246 EQLOEENFRL-----ESGREDELRCAELERVAELQH-----RNQ 281
Db 557 YQAOVEMRLMNMNLEEDIVSARRSDLYESSELRESLAAEEFKRKATECOHLLKAKQ 616
QY 282 ALTSLAQEA-----QALKDEMDERLROSSERAG--OLEATLTSCRRRLGELRLRRQV 331
Db 617 GKPEVGEYAKLEKINAEQOLK--IQELQEKLEAKAERARELEKLNRRDSEGIKKL 674
QY 332 RQLEERNAGHAETRQLEDELRRAGSLRAQLEAQRQVQEL-----372
Db 675 VEAEERRHSLNKKVRLKLETMRERENRDKDDIOTKSQOIQOMADKILEEKEHREAOVSAQ 734
QY 373 ----QGOROEEMK-----AEKWLFECEKNLEKESVTKKERLLAERDSL 414
Db 735 HLEVHLKQKEQHYEEKIKVLDNOIKKDLADKETLE--NMQRHEEAHEKGLTSSQKAM 792
QY 415 REANE-----ELRCQLOPRLGTQADPSLDPTSTPVNLAAILPAELRETLRLRLQLEN 468
Db 793 INAMDKIRSLQRIVELSEANKLAANSSL---FTQNNKAQEMISELRLQOKFYLETQA 849
QY 469 KRLCROEAADRRQOEELQ--HLEDANRARGHLETOHRL-----NOQQLSELRQAQVDEL 520
Db 850 GKL---EAQNRKLEEQLEKISHQDSDKNRL--LELETRRLREVSLEHEEQKLEKRLQTEL 905
QY 521 OKALOQ-----GGKTEDAISIL-----LKRK-----542
Db 906 QLSQERESQLTALQARAALAESQLQAKTELEETTAEEETQALTAHRETOIRKFDAL 965
QY 543 -----LEEHLQKHEADLELQKREYIEELEPPTDSSTARREEELQHNKQKDAOL 593
Db 966 RNSCTVITDLEEQNLQNTEDNAELNNQNFY---LSKOLDEASGANDEIVQ-----L 1013

APPLICANT: Kapeller-Libermann, Rosana
TITLE OF INVENTION: 13245, A No. US20020160483A1 Human Myotonic Dystrophy Type P
FILE REFERENCE: Kinase and Uses Therefor
CURRENT APPLICATION NUMBER: US/10/017,216
PRIOR FILING DATE: 2001-10-23
PRIOR APPLICATION NUMBER: US 60/242,429
PRIOR FILING DATE: 2000-10-23
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2053
TYPE: PRT
ORGANISM: Homo sapiens
US-10-017-216-2

Query Match 8.0%; Score 290; DB 9; Length 2053;
Best Local Similarity 20.2%; Pred. No. 1.1e-08;
Matches 162; Conservative 134; Mismatches 262; Indels 242; Gaps 28;

QY 92 SEEHLPDV-----SLIGESDPAELGKLLQLVL-----GCAISCEKKQ 129
Db 410 SGEELPFVGSYKALGILGRSESVGLDSPAKTSMKKLLIKSKELQDSQDKCKME 469
QY 130 DHQIRIMTLEESVQHVMEAIQELMTKDPDSLSPETGYNFDQSRRYYFLSEAEGBE 189
Db 470 QEMTRLHRRVSEVAVLSQKEVELKASQTSRSLLEQDLATYITE-----513
QY 190 LQORCLDLERQL-----MLLSEEQSLAQENAGLRNMRGPEGETGLTAKKLLQLSQ 245
Db 514 ----CSSLRSLQARMEVSQEDDKALQLLDHIREQ-----SRKLOEIKQE 556
QY 246 EQLOEENFRL-----ESGREDELRCAELERVAELQH-----RNQ 281
Db 557 YQAOVEMRLMNMNLEEDIVSARRSDLYESSELRESLAAEEFKRKATECOHLLKAKQ 616
QY 282 ALTSLAQEA-----QALKDEMDERLROSSERAG--OLEATLTSCRRRLGELRLRRQV 331
Db 617 GKPEVGEYAKLEKINAEQOLK--IQELQEKLEAKAERARELEKLNRRDSEGIKKL 674
QY 332 RQLEERNAGHAETRQLEDELRRAGSLRAQLEAQRQVQEL-----372
Db 675 VEAEERRHSLNKKVRLKLETMRERENRDKDDIOTKSQOIQOMADKILEEKEHREAOVSAQ 734
QY 373 ----QGOROEEMK-----AEKWLFECEKNLEKESVTKKERLLAERDSL 414
Db 735 HLEVHLKQKEQHYEEKIKVLDNOIKKDLADKETLE--NMQRHEEAHEKGLTSSQKAM 792
QY 415 REANE-----ELRCQLOPRLGTQADPSLDPTSTPVNLAAILPAELRETLRLRLQLEN 468
Db 793 INAMDKIRSLQRIVELSEANKLAANSSL---FTQNNKAQEMISELRLQOKFYLETQA 849
QY 469 KRLCROEAADRRQOEELQ--HLEDANRARGHLETOHRL-----NOQQLSELRQAQVDEL 520
Db 850 GKL---EAQNRKLEEQLEKISHQDSDKNRL--LELETRRLREVSLEHEEQKLEKRLQTEL 905
QY 521 OKALOQ-----GGKTEDAISIL-----LKRK-----542
Db 906 QLSQERESQLTALQARAALAESQLQAKTELEETTAEEETQALTAHRETOIRKFDAL 965
QY 543 -----LEEHLQKHEADLELQKREYIEELEPPTDSSTARREEELQHNKQKDAOL 593
Db 966 RNSCTVITDLEEQNLQNTEDNAELNNQNFY---LSKOLDEASGANDEIVQ-----L 1013
```

```
QY 594 RAMEERYRYVDKARVMY---OTMEPKQRPAAAGP---ELHSLRTOLRER-----638
Db 1014 RSEVDHLRREITEREMOLTSQKOTMEALKTTCTMLEEQVNDLALNDELLEKERQWEAWR 1073
QY 639 -----DVRIRHLEMDFKRSRQOEKILLISAWYNGMALQORAGERRAPAHQAQ 688
Db 1074 SVLGDEKSFQECRVRELQRLMDTEKQSRARADQRTESRQVVELAVKEHKAELA---LQ 1130
QY 689 SFLAQORLATNSRRGPIGLRL 708
Db 1131 QALKEQKLAESLSDKLNDL 1150

RESULT 9
US-10-028-946-2
; Sequence 2, Application US/10028946
; Patent No. US20020123622A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; APPLICANT: Fiddle, Carl Johan
; TITLE OF INVENTION: No. US20020123622A1el Human Kinases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0289-USA
; CURRENT APPLICATION NUMBER: US/10/028,946
; CURRENT FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 60/258,335
; PRIOR FILING DATE: 2000-12-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2054
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-028-946-2

Query Match 8.0%; Score 290; DB 12; Length 2054;
Best Local Similarity 19.8%; Pred. No. 1.1e-08;
Matches 162; Conservative 137; Mismatches 258; Indels 260; Gaps 28;

QY 92 SEELPDV-----SLIGFSDPAELGKLQLVL-----GCAISCEKKQ 129
Db 410 SGEELPVGFYSKALGILGRSESVGLSDSPAKTSSMEKKLLIKSKELQDSODKCHKME 469
QY 130 DHTQRTIMLEESVQHVVMIAIOLMTKDPDLSLSPETYGNFDSQSRYYFLSEAEEGDE 189
Db 470 QEMTRLHRRYSEVAVLSQKEVELKASETQSLLEQDLATYITE-----513
QY 190 LQORCLDLERQL---MLLSEKQSLAQENAGLRMRGRPEGETPGLTAKKLLQLLQSQL 245
Db 514 ---CSSLKRSLQARMEVSQEDDKALQLLHDIREQ-----SRKLOEIKQE 556
QY 246 EQLOEENFRL-----ESGREDELRCAELEREVAELQH-----RNQ 281
Db 557 YQAQVEEMRLMNQLEEDLVSAARRSDLYESELRESLAEEFKRKATECOHLLKAKQ 616
QY 282 ALTSIAQEA-----QALKDEME-----LRQSSERAGLEATL 314
Db 617 GKPEVGAYAKLEKINAEQOLKIOELQEKLEKAVKASTEATELLQNIQAQERA---EREL 673
QY 315 TSCRRRLGELRELRRQVROLEERNAGHAERTROLEDELRRAGSLRAQELQRRQVQL-- 372
Db 674 EKLQNRQSSSEGIRKKLVEAEERHSLNKKVRLTWERRENLUKDDIOTKSOQIQOMAD 733
QY 373 -----QGOROEAMK-----AEKWLFECCNLPEKY 397
Db 734 KILELEBKHEAQAQVSAQHLVHLKQEHQVKEIKVLDNQIKDLADKETELE--NMQRH 791
QY 398 ESVTKERLLAERDSUREANE-----ELRCAQOLPRGTQADPPSLDPTSTPVDNLAAE 451
Db 792 EEEAHGKILSEQKAMINAMDSKIRSLQORIVELSEANKLAANSSL---FTQNMKAQE 848
QY 452 ILPAELRETLRLQLENKRLCRQEAADROEOELQR--HLEDANRARGHLETOHRL-----505
```

```
Db 849 EMISELRQOQFYLETQAGKL---EAQNRKLEOELEKISHODSHDKNRL-LELETRLREVS 904
QY 506 --NQOQLSELRAQVEDLQALQEQ-----CGKTEDAI 535
Db 905 LEHEEQKLELKRQTLTQLSLOQRESQLTALQARAALSQLRQAKTELETTAAEAEI 964
QY 536 SIL-----LKRK-----LEEHLQKLHEADLELQKREYIELEPPPTSSSTA 576
Db 965 QALTARDEIQRFDAIRNSCTVITDLEOLNQLTDEQALNQNQFY---LSQOLDEASG 1021
QY 577 RRIEELQHNLOKKDADLRAEMEYRYRYVDKARVMY---OTMEPKQRPAAAGP---ELH 629
Db 1022 ANDEIVQ-----LRSEVDHLRREITEREMOLTSQKOTMEALKTTCTMLEEQVNDLE 1072
QY 630 SLRTOLRER-----DVRIRHLEMDFKRSRQOEKILLISAWYNGM 671
Db 1073 ALNDELLEKERQWEAWRSVLGDEKSFQECRVRELQRLMDTEKQSRARADQRTESRQV 1132
QY 672 MALQORAGERRAPAHQAQSLAQORLATNSRRGPIGLRL 708
Db 1133 LAVKEHKAELA---LQOALKEQKLAESLSDKLNDL 1166

RESULT 10
US-10-017-216-4
; Sequence 4, Application US/10017216
; Patent No. US20020160483A1
; GENERAL INFORMATION:
; APPLICANT: KAPPELLER-LIBERMANN, Rosana
; TITLE OF INVENTION: 13245, A No. US20020160483A1el Human Myotonic Dystrophy Type P
; TITLE OF INVENTION: Kinase and Uses Therefor
; FILE REFERENCE: 10147-5701
; CURRENT APPLICATION NUMBER: US/10/017,216
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,429
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2055
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-017-216-4

Query Match 8.0%; Score 290; DB 9; Length 2055;
Best Local Similarity 20.0%; Pred. No. 1.1e-08;
Matches 165; Conservative 132; Mismatches 266; Indels 262; Gaps 28;

QY 86 VLAPV---SEELPDV-----SLIGFSDPAELGKLQLVL-----GC 121
Db 401 VPAEPFLAFSGEELPFVGFYSKALGYLGRSESVVSSLDSPAKVSSMEKKLLIKSKELQDS 460
QY 122 AISCEKKQDHTQRTIMLEESVQHVVMIAIOLMTKDPDLSLSPETYGNFDSQSRYYFLS 181
Db 461 QDKCKMEQEWTRLHRRYSEVAVLSQKEVELKASETQSLLEQDLATYITE-----512
QY 182 EEAEQDELOQRCLDLERQL---MLLSEKQSLAQENAGLRMRGRPEGETPGLTAKK 237
Db 513 -----CSSLKRSLQARMEVSQEDDKALQLLHDIREQ-----SRK 547
QY 238 LLLQSOLEQLQENFRL-----ESGREDELRCAELEREVAELQH 278
Db 548 LOETKEQYQAQVEEMRLMNQLEEDLVSAARRSDLYESELRESLAEEFKRKANECQH 607
QY 279 R-----NQALTSIAQEAQALKDEME-----LRQSSER 306
Db 608 KLMKAKDQGPPEYGEYSKLEKINAEQOLKIOELQEKLEKAVKASTEATELLQNIQAQER 667
QY 307 AGOLEATLTSCRRRLGELRELRRQVROLEERNAGHAERTROLEDELRRAGSLRAQLEAQR 366
Db 668 A---ERELEKLNHNPDESSEGIKKKLVAEERHSLNKKVRLTWERRENLUKDDIOTKS 724
QY 367 RQVQL-----QGOROEAMK-----AEKWLFE 389
```

```

Db 725 EOLIOQWADKILLEEKREAOVSAGHLEVHLKOKEOHYEEKIKVLDNOJKKLADRESLE 784
QY 390 CRNLEKYSVYKEREKRLAERDSLREANE-----ELCAQOLQPGGLQADPSLPTST 443
Db 785 --NMQRHEEHEKELISEQKAMINADSKIRSLQRIIVELSEANKLAANSL--PT 839
QY 444 PYDNLAEILPRLRETLRLIOENKRLCROEAADREBOELOR--HLEDANKARGLLET 501
Db 840 QRMKRAQOEKISELROQKTYLETQAGKL--EKONKRLSEOLEKISHODSKSRL-TEL 895
QY 502 QHRL-----NOOQSELRAVOEDLOKALOEO----- 527
Db 896 EURLRVSLEHEQOKELKRLQLEIQLSLQERESQTLQAARAALESOLROAKTELEET 955
QY 528 GGTEDATIL-----LKR-----LEHLQKLEADLELQKREYEELE 568
Db 956 TAEAEELIALTAHDEIQRKFDALNSTCTVTDLEEQNLQLEDAEALNNQFY--LS 1012
QY 569 PPTDSSTARIELOHNLQKADLRAEMEERYRKYDKARMY--OTMEPRQAPAGA 624
Db 1013 KOLDEASGANDIYO-----LRSEVDHLRREITEREMQLTQKOTWEALKTCTYML 1063
QY 625 PE--ELMSLRQLRER-----DVRIRHLEMPDEKSRQSEOEKIL 663
Db 1064 EEOVDLEALNDELLEKEREQWEAMRSVLGDEKSQFECRVRELOQRLDTEKQSRARADQRI 1123
QY 664 ISAWYNGMALQORAGEERAPAAQSFLAQORLATNSRRGPGRL 708
Db 1124 TESROYVELAVEKHAELIA--LOALKEOKIKAESLSDKLMDL 1165

RESULT 11
US-09-884-001-19
; Sequence 19, Application US/09884001
; Publication No. US20020182656a1
; GENERAL INFORMATION:
; APPLICANT: Bird, Timothy A.
; APPLICANT: Peschon, Jacques J.
; APPLICANT: Sims, John E.
; APPLICANT: Vitica, G. Duke
; APPLICANT: Willis, Cynthia R.
; TITLE OF INVENTION: Methods for Regulating Vascularization Using GGF.
; TITLE OF INVENTION: Containing NEK-Like Kinase (GNK)
; FILE REFERENCE: ImmuneX GNK/GNK PCT
; CURRENT APPLICATION NUMBER: US/09/884,001
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/113,003
; PRIOR FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-001-19

Query Match 7.8%; Score 284.5; DB 9; Length 868;
Best Local Similarity 23.8%; Pred. No. 8.9e-09;
Matches 139; Conservative 106; Mismatches 189; Indels 151; Gaps 24;

QY 180 LSPAEAEQ-----ELQORCLDLERMLLS-----EKKOSLAQENAGLRERMGRE----- 226
Db 36 LSSSRKQOEAAATTOQLQOEAKROEVLARAVOEKALVREKAALEVLRLQAVERRDRL 95
QY 227 GEGTPELTKKLLLOSQLOEQENFRLE--SGREDERL-----RCAEL- 269
Db 96 AEGLOGLSSAK-ELLESSLFEAQONSVEYVTKGLEVQIQVTAQKAVYQGVRLKLE 154
QY 270 ---EREVAELQRNOALJSLAQPAQALKDEMDELROSSFERAGOLEATVTSRRRLGELRE 326
Db 155 LPTERSQAE-QERDAALAROLAQAEQGTALREQOKAAHEK-----EVNQ 197

```

```

QY 327 LRROVLEERNAGHAERTROLEDELRRAGSLRAQLEAORQVOELQGOEAMAEK 386
Db 198 LRKWE--KERSHQOEALKALESLEREKMELEMLKEQOTEALQAOERREERTQAES 255
QY 387 LECRNLEEKYESVYKEREKRLAERDSLREANEELRCALOQPGGLQADPSLPTSTPYD 446
Db 256 LCOMQLETER-----ERVSILET-----LLQTKELADASQOLE 289
QY 447 NLAELPRLRETL-----LRLIOENKRLCROEADRER-----QEE-----LQ 486
Db 290 RLQDMKVOKLQOEETTGILQTOLEAQORELKEAAROHDLAALQEESSSLQDQMDQ 349
QY 487 RHLEDANRANHGLETOHRLNOOQSELRAVOED--LOKALQOGKTDATSIILKRL 543
Db 350 KOVEDLKSQIVADDSQORLQVEQVOKLEETOENRQKLELE--KASLTJSLMKR--- 404
QY 544 EELQKRLHAD-----LELQKREYEELEPPTDSSTARIELOHNLQKADLRAEMEER 599
Db 405 EORLVLQENDSIRQOELSALRODMQEAQGEQKELSA-QMELLROGVKREKDFLAQEA 463
QY 600 YRRYVKAAMVQMTPEPKORPAAQAPPELSLRTOLRENDVRIRHLEMPDEKSRQEOE 659
Db 464 LLEELASHITEO-----QLRASLMAQEKAKAOLQLRLRSTESOLE-- 504
QY 660 EKLLISAWYNGMALQORAGEERAPAAQSFLA-----QORLAT 698
Db 505 -----ALAEQOPGNO--AAQOALASLYSALQALGS 535

RESULT 12
US-10-017-216-5
; Sequence 5, Application US/10017216
; Patent No. US20020160483a1
; GENERAL INFORMATION:
; APPLICANT: KAPEDLER-LIBERMANN, Rosana
; TITLE OF INVENTION: 13245, A No. US20020160483a1el Human Myotonic Dystrophy Type P
; TITLE OF INVENTION: Kinase and Uses Therefor
; FILE REFERENCE: 10147-5701
; CURRENT APPLICATION NUMBER: US/10/017,216
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,429
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1641
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-017-216-5

Query Match 7.8%; Score 282; DB 9; Length 1641;
Best Local Similarity 20.9%; Pred. No. 2.5e-08;
Matches 162; Conservative 130; Mismatches 264; Indels 220; Gaps 28;

QY 92 SEEHLPDV-----SLIGFESDPALCKLQIVL-----GCALISEKKQ 129
Db 37 SGEELPVGFSYKALGYGRSESVSSLDSPKAVSMERKLLIKREKIQDSODKCHME 96
QY 130 DHIQRIIMTEESVOHYVMAIOELMTKPTDLSLPTYNFDSQSRYYFLSEAEQGE 189
Db 97 QEMTRLHRRYSEVAVLSQKEVELKASETQRSLLBQDLATYIRE----- 140
QY 190 LQORCLDLEROL-----MLSEKQSLAQENAGLRERMGREGGTPGLTAKKLLLOSQ 245
Db 141 ---CSSLKRSLQOARMEVSOEDBKALQLLHDHREO-----SRKQLEIKQOE 183
QY 246 EOLQENFR-----ESGREDERLCAELEREVALEOH-----RNO 281
Db 184 YQAVQVEEMRLMANQLEEDLYSARRRSDLYSELRERSLAAEFKRRANCOHRLMAKQO 243
QY 282 ALTSLAQEAQALNDEMDELROSSERAGOLEATVTSRRRLGELRELRROVROLEERNACH 341
Db 244 GRPEVGEYSLEKINAEQQLKIOELQEKLEKAVKAST-----EATELLQNIQAKKEAENE 299

```

342 AERTQLED-----ELRRAGSL-----RAQLEAQRORVO-ELQGOROEAMK----- 382  
Db 300 LEKUNREDSSEGKKKVUVAEELEEKHREAQVSAQHLEHLKQEQHYEKIKVLNOI 359  
383 ---AEKWLFCRNLLEEKYEVTEKERLLAERDSLEANE-----ELRCAQLQPRGLT 432  
Db 360 KKLADKESLE--NMQRHEEAEHKGKILSEQKAMINAMDSKIRSLERQIVELSEANKL 417  
383 QADPSLDPTSTPDVNLAAEIPAELETLRLQLENKRLKRCQEAADRRERQBELQR--HLE 490  
Db 418 AANGSL---FTORNKAQEMISELROQKFYLETOAGKL---EAQNRKLEQOLEKISHOD 471  
383 DANRAGHLETOHRL-----NQOQLESLRAQVEDLQALQEO----- 527  
Db 472 HSDKSLR-LELETLRLVRELSHEEQLKRLQTLQSLQERESQTLQAQARALESQ 530  
383 ---GGKTEDAISIL-----LKKK-----LEHLKQLHEADLEL 557  
Db 531 LQAKTELEETAEAEIEIQALTAHRDEIQKFDALRNSCVITDLEQLQALQEDNAEL 590  
383 QKREYIEELPPTSDSTARRIEELQHLQKDDADLRAMEERYRYVDKARVM-----OT 613  
Db 591 NNQNFY---LSQLEDEASGANDIIVQ-----LRSEVDHLRREITEREMQLTSQKT 638  
383 MEKQORPAGAPP---ELHSLRTOLRER-----DVRIRHLEMDPEKS 652  
Db 639 MEALTKTCTMLEEQVLDLEALNDELLEKERQEWAEWRSVLGDEKSFQFCEVRLEQLMDTE 698  
383 RSQRQEEKLISANYNMGALQORAGERAPAHQAQSLAQORLATNSRRGPLGRL 708  
Db 699 KOSRARADQRTESRQVVELAVKEHKAELIA---LQALKKEQKLKAESLSDKLNDL 751

## RESULT 13

US-09-815-242-11828  
Sequence 11828, Application US/09815242  
Patent No. US20020061569A1

## GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes  
FILE REFERENCE: ELTRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11828  
LENGTH: 1162  
TYPE: PR1  
ORGANISM: Pseudomonas aeruginosa

US-09-815-242-11828

Query Match 7.6%; Score 277; DB 10; Length 1162;  
Best Local Similarity 23.8%; Pred. No. 3.2e-08;  
Matches 174; Conservative 113; Mismatches 248; Indels 196; Gaps 34;

QY 26 ASPODLSGLA--VAYVINOIDPSWFNEMLQ-----GISEDGPNWKLKVSUKVLR 77  
Db 353 APEQLSAAAEAGIALEQAEOGM--QAWOOQWDAFNQQAEPRAEQAEVQOQSIOHLEQ 410  
QY 78 SLVYSODVLAHPVSEHL-----PDVSLIGFSDPAELGKL-----LQVLGCAI 123  
Db 411 SL-ERLQD-----RERRLOERQGLAADPEDAAILENEQVAIAELALELQ- 457  
QY 124 SCEKKODHQRIMTLESYQHV-----MEAIQELMTKOTPDLSLSP 164  
Db 458 ---EQGQAELEQLRQLEQLOQAEEQAQGELOLNGRIASLEALQQAALDPQGALEW 514  
QY 165 ETYGNFQSRRY-----YFLSEAEDEGLQORCLD-----LEROLMLSE 206  
Db 515 LREQGLEQRPRLAEGLRVEPGWELAVETVLGADLQAVLLDFGLAGFGKGLRLILSP 574  
QY 207 EKQSLAQENAGLRMRGPCEGTPGLTAKKLLLLQSOLEQLOQENPELESREDERLRC 266  
Db 575 AR-GAATAAGSLDKV-RADADLSPLARVKPV---ETLEQALQARGALDDG---ESLIS 626  
QY 267 AE---LEREVAELQHRNOA---LTSLAQEAQALKDEMDELQSSERAGOLEATLTSCRRR 320  
Db 627 RDGYWVGHFLEVRSDENAGOMLARAQALEALQERREAL---ETRVAGEERLAAARDE 683  
QY 321 LGEIRLRRQVRQLEERNAGHAETRQLEDELRLRAGSLRAQLEAQRORVOLEQOOR--- 376  
Db 684 QRELEGAREQVR-----RQVQEEGRRHGELKAQLSAQAKVEQLVLRRLD 730  
QY 377 ---QEEAMKAEKWLFEGR-NLEEKYEV---TKEKERLLAERDSLEANEELR--- 422  
Db 731 BEVAELAEQRALEQEQ-LSEARLTQEAALDQWALDTERRESLLAERDALRDLRIRODA 789  
QY 423 ---CAQLQPRGLTQADPSLDPTSTPYD-----NLAAEILPAELRETL 461  
Db 790 RTHKDHQHLAVRVGSLKAQHNSTQALERLDQQSARLNERCEQLNLNLEGAAPLEE-- 847  
QY 462 LRLQLENKRLCQEAADRRERQEEELQRLHLEDANRARGHLETOHRLNQQLSELRAQVED-- 519  
Db 848 LRMKLEELERRMAVEDLQKARLA--LEDADRELREVEKRRGQAEQOQSLLRGOLEQOR 905  
QY 520 ---LQALQEQGQKTEDAISI-----LKKRLEEHLQKL----- 550  
Db 906 LEWQGLVVRKALQEQ--LAEDGYDLHTVLNPLDASERDWEERLESLSAARIQRLGPIIN 963  
QY 551 ---HEADLELQKREYTELEPPTSDSTARRIEELQHNLOKKDADLRAMEERYRYVDK 606  
Db 964 LAIEEYQOQSERRK-YLDS-----QNDDLAEALETLENVIRKIDRETR---NRKFETDQ 1015  
QY 607 ARVMQTMPEK 617  
Db 1016 INAGLQALFPK 1026

## RESULT 14

US-10-017-216-6  
Sequence 6, Application US/10017216  
Patent No. US20020160483A1

## GENERAL INFORMATION:

APPLICANT: KAPELLER-LIBERMANN, Rosana  
TITLE OF INVENTION: 13245, A No. US20020160483A1 Human Myotonic Dystrophy Type P  
FILE REFERENCE: Kinase and Uses Therefor  
FILE REFERENCE: 10147-5701  
CURRENT APPLICATION NUMBER: US/10/017,216  
CURRENT FILING DATE: 2001-10-23  
PRIOR APPLICATION NUMBER: US 60/242,429  
PRIOR FILING DATE: 2000-10-23  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 6  
LENGTH: 1597  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-017-216-6

Query Match 7.6%; Score 276; DB 9; Length 1597;  
Best Local Similarity 20.5%; Pred. No. 5.2e-08;  
Matches 145; Conservative 134; Mismatches 258; Indels 170; Gaps 23;

QY 128 KOD---HIQIMTLESVOHVYMEATQELMTKDTDPSPISPEYGNFDSQSRYYFLSEEA 184  
DB 45 EQDLATYITCSLSKRSQARMEVSQE-----DDKALQLLHDIREOSKRLQEIKEQE 97  
QY 185 EGDELOQRCCLDERQLMLLSEKQSLAENAGLRMRGPEGPGTAKKLILLOQ 244  
DB 98 YQAVQEMRLMNQLEEDLVARRRDLSELSRESRLAAEFKRNKANECHKLMAKQO 157  
QY 245 LEQLOQENFLESGBEDRLCAELREYAEVLAQRNQAALSLAQAQALKDENDLRQSS 304  
DB 158 GRPEVGEYSKLEKINAQOLK-----IOLEQKLEKAVKASTATEL---LQNTQAK 207  
QY 305 ERAGOLEATLTSCRRRLGELRLRQVROLEARNAGHAERTQLEDELRRAGSLRAQLEA 364  
DB 208 ERA---ERELEKLNHRNEDSSEGIKKKLVAEERRHSLENKVKRLTMRERENKDDIQT 264  
QY 365 QRRQVQEL-----QGRQOEAMK-----AKWL 387  
DB 265 KSEIQOQMAKILEEKEHREAQVSAQHLVHLKQEOHYEIKIKVLDNQIKDLADKES 324  
QY 388 FEQRNLEEKYESTKEKRLAERDSREANE-----ELRCAQLOPRLGTQADPSLDPT 441  
DB 325 LE---NMQRHEEAEHEKGLSEQRAMINAMDSKIRSLQRIVELSEANKLAANSL--- 379  
QY 442 STPVDNLAETIPAELETLRLQLNKLRCQEAADRERQELQ---HLEDANRAHGL 499  
DB 380 FTQRNKAQEMISELROCKFYLETQAGKL---EAQRNLEQLEKISHODSDKSL-L 435  
QY 500 EQHRL-----NQOQLSELRAQVEDLOKALQEO-----527  
DB 436 ELETRLVSLHEEOKLEKQLTLEQLSLORESQLTALQARAALLESQLRQAQTELE 495  
QY 528 ---GGKTEDAISL---LKRK-----LEEHQKLEADLEQKREYIEE 566  
DB 496 ETTAEAEIEQALTAHRDIQKFDALRNSCIVITDLEEQNLQTEDNAELNNQNY--- 552  
QY 567 LEPPDTSSTARIEELQHLNKLQKADDLRAWEERYRYVDKARMV-----OTWEPKORPAA 622  
DB 553 LSKQLEAGSGANDEIVQ-----LRSEVDHLRREITEREMQLTQKOTMEALKTTCT 603  
QY 623 GAPP---ELHSLRTQLRER-----DVRIRHLEMDFKRSRQREKEK 661  
DB 604 MLEEQVLDLEALNDELLEKERQWEAWRSVLGDEKSFQFECVRELQRLMLDTEKOSRARQ 663  
QY 662 LLISAWYNGMALQORAGGERAPAHQSFLOQRATNRRGRPLGL 708  
DB 664 RITESQVVELAVKEHKAELIA---LQALKQKLAESDKLNDL 707

RESULT 15  
US-10-108-605-217-  
Sequence 217, Application US/10108605  
Patent No. US20020160934A1  
GENERAL INFORMATION:  
APPLICANT: Broadus, Julie  
APPLICANT: Stam, Lynn  
APPLICANT: Bachmann, Jane  
APPLICANT: Kamdar, Kim  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE  
FILE REFERENCE: 31133B  
CURRENT APPLICATION NUMBER: US/10/108,605  
CURRENT FILING DATE: 2002-03-27

PRIOR APPLICATION NUMBER: US 09/761,142  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/176,418  
PRIOR FILING DATE: 2000-01-14  
NUMBER OF SEQ ID NOS: 361  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 217  
LENGTH: 879  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-10-108-605-217

Query Match 7.4%; Score 270.5; DB 9; Length 879;  
Best Local Similarity 21.7%; Pred. No. 5.5e-08;  
Matches 150; Conservative 136; Mismatches 258; Indels 147; Gaps 29;

QY 40 VLNQIDPSWFNAWLOGISEDPPNWKLVNLMKVLRSLEYSDVLAH---PVSEHL- 96  
DB 160 LLSQIES--YNEKEI--VSEKHISKLEVISSELNVKIEELNRTVIDISSHRSRLSQENIE 215  
QY 97 --PDVSLIGFSDPAELGKLQLVLGCAISCCKKQDHQIORTIMTLESVOHVYME---AIOE 152  
DB 216 LTKDVQDLKVQDVTVSFSK--SQVISQLEDARRRLEDEDRRRSLLESSLHQVETELDSVRN 274  
QY 153 LMTKDPDSLSPE---TYGNFDSQSRYYFLSEAEEGDELOQRCCLDERQLMLLSEEQ 209  
DB 275 QLEESEARIDLERQLVKANADATSMQNKWSEVAARAEVEEIRKRYQVRITELBEHIE 334  
QY 210 SLAQENAGLRMRGPEGPGTAKKLIL-----LQSOLEQOENFRLSEG 258  
DB 335 SLIVKYNLEKMKTRLASE---VEVLIDLEKSNNSCRELTKSVNTEKHNVELKS- 387  
QY 259 REDERURCAELREYAEVLAQRNQAALSLAQAQALKDENDLRQSSERAG---QLEATL 314  
DB 388 RLDETILYETSQR---DLKNHADLVRTVHDLKVDKNNNNQLTRENKLGDDOLHEAKGAI 445  
QY 315 TSCRRRLGELR-ELRQVQLERNAHAERTQLEDELRRAGSLRAQLEAQRQVQELQ 373  
DB 446 NELNRLHELELRLENERDELTAAYE-----AEAG-----RKAEEQ 486  
QY 374 GORQEEAMKAELWFEKRNLEEKYESTKEKRLAERDSREANEELRCAQLOPRLGTQ 433  
DB 487 GQR-----LAADFNOYRHAERRLAEKO-----EEIEAIRKQ-----518  
QY 434 ADPSLDPTSTPVDNLAETIPAELETLRLQLNKLRCQEAADRERQELQRLHEDAN 493  
DB 519 -----TSIEIEQLNARVIEAE---TRLKTEVTRIKKKL---QIQITELEMSLDVAN 563  
QY 494 RARHGLETHRLNQOQLSELRAQVEDLOKALQEOGGKTEDAISILKR-----KLEE-- 545  
DB 564 KTNIDLOKVIKKSQULTELOAHYEDVQROLQ-----ATLDQYAVAAORRLAGLNGELEEV 619  
QY 546 -HLQKLHEA---DLELQKREYIELEPPDTS--STARRIE---LQHNQKQKADLR 594  
DB 620 SHLDSANRAKRVLEQYEEASFINELTTANVSLVSKLQELSVASDVSEVSKELR 679  
QY 595 AMEERYR-YVDKARMVMQTMPEKQRPAA-----AGAPPELHSLRTQLRE-----637  
DB 680 ISDERYQKVQVELKHVVEQVHEEOERIVKLETIKKSLEVEKVNLSIRLEEVELNAVAGSK 739  
QY 638 -----RDVIRHLEMDFKRSRQREKEKLL 663  
DB 740 RIISKLEARIRDELEEBEKKRRHAETIKIL 770

Search completed: March 4, 2003, 15:07:24  
Job time: 31.2109 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:50:41 ; Search time 14.9505 Seconds  
(without alignments)  
1559.740 Million cell updates/sec

Title: US-09-763-902B-6  
Perfect score: 936  
Sequence: 1 MGILLGLLLGLHVTDTYGR.....PLKATSTVQKSWDWTMDMG 175

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_101002:.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Score | Length | DB ID | Description                  |
|------------|-------------|-------|--------|-------|------------------------------|
| 1          | 936         | 100.0 | 175    | 21    | AA1982322 Human protein tran |
| 2          | 936         | 100.0 | 306    | 20    | AA130814 Human secreted pro  |
| 3          | 870         | 92.9  | 321    | 20    | AA141691 Human PRO 362 prot  |
| 4          | 870         | 92.9  | 321    | 20    | AA123322 Amino acid sequenc  |
| 5          | 870         | 92.9  | 321    | 21    | AA144247 Human PRO362 (UNQ3  |
| 6          | 870         | 92.9  | 321    | 21    | AA124047 Human PRO362 prote  |
| 7          | 870         | 92.9  | 321    | 21    | AA133429 Human PRO362 prote  |
| 8          | 870         | 92.9  | 321    | 21    | AA195346 Amino acid sequenc  |
| 9          | 870         | 92.9  | 399    | 21    | AA19396 Human polypeptide,   |
| 10         | 870         | 92.9  | 399    | 22    | AA193874                     |

|    |       |      |      |    |                             |
|----|-------|------|------|----|-----------------------------|
| 11 | 870   | 92.9 | 399  | 22 | AAE04208 Human gene 11 enco |
| 12 | 870   | 92.9 | 399  | 22 | AAE04230 Human gene 11 enco |
| 13 | 870   | 92.9 | 399  | 23 | ABG64487 Human albumin fusi |
| 14 | 870   | 92.9 | 399  | 23 | ABG64488 Human albumin fusi |
| 15 | 807   | 86.2 | 386  | 22 | AAE04290 Human gene 11 enco |
| 16 | 733.5 | 78.4 | 184  | 22 | AA193588 Human polypeptide, |
| 17 | 492   | 52.6 | 93   | 22 | AAE10596 Human macrophage-e |
| 18 | 130.5 | 13.9 | 360  | 20 | AA113381 Amino acid sequenc |
| 19 | 130.5 | 13.9 | 360  | 22 | AAU23037 Human PRO polypept |
| 20 | 130.5 | 13.9 | 360  | 22 | AA138965 Human polypeptide  |
| 21 | 130.5 | 13.9 | 360  | 22 | AA180249 Human PRO271 prote |
| 22 | 130.5 | 13.9 | 172  | 22 | AA127244 Human EXMAD-22 SEQ |
| 23 | 127.5 | 13.6 | 289  | 21 | AA193912 A human hyaluronan |
| 24 | 109   | 11.6 | 434  | 22 | ABG02117 Novel human diagno |
| 25 | 109   | 11.6 | 737  | 22 | ABG02019 Novel human diagno |
| 26 | 109   | 11.6 | 5635 | 23 | ABF60991 Human LP polypepti |
| 27 | 107   | 11.4 | 402  | 23 | AAU76528 Novel human secret |
| 28 | 107   | 11.4 | 402  | 23 | AAU09883 Human A33 antigen. |
| 29 | 102.5 | 11.0 | 319  | 18 | AAW14146 Amino acid sequenc |
| 30 | 102.5 | 11.0 | 319  | 20 | AA123323 Human A33 protein  |
| 31 | 102.5 | 11.0 | 319  | 22 | AA157083 Rat proteoglycan 1 |
| 32 | 102.5 | 11.0 | 339  | 21 | AAE23290 Human nectin-3beta |
| 33 | 102   | 10.9 | 387  | 23 | AAE23289 Human nectin-3alph |
| 34 | 102   | 10.9 | 426  | 23 | AAE23299 Human nectin-3gamm |
| 35 | 102   | 10.9 | 437  | 23 | AAE23284 Human deleted nect |
| 36 | 102   | 10.9 | 504  | 23 | AAE23285 Mouse nectin-3-hum |
| 37 | 102   | 10.9 | 510  | 23 | AAE23286 Human nectin-3beta |
| 38 | 102   | 10.9 | 510  | 23 | AAE23281 Human deleted nect |
| 39 | 102   | 10.9 | 542  | 23 | AAE23282 Mouse nectin-3-hum |
| 40 | 102   | 10.9 | 549  | 23 | AAE23283 Human nectin-3alph |
| 41 | 102   | 10.9 | 549  | 23 | AAE23283 Human polypeptide  |
| 42 | 102   | 10.9 | 555  | 22 | AA139143 Human polypeptide  |
| 43 | 102   | 10.9 | 559  | 22 | AA140929 Human nectin-3beta |
| 44 | 102   | 10.9 | 595  | 23 | AAE23288 Human nectin-3alph |
| 45 | 102   | 10.9 | 634  | 23 | AAE23287                    |

ALIGNMENTS

RESULT 1  
AA1982322  
ID AA1982322 standard; Protein; 175 AA.  
AC AA1982322;  
DT 19-JUN-2000 (first entry)  
XX Human protein transport molecule (PTAM) SEQ ID NO:6.  
XX Human; protein transport molecule; PTAM; diagnosis; cytostatic;  
KW antithalritic; antidiabetic; immunosuppressant; antiarteriosclerotic;  
KW antiallergic; antidiabetic; antilipemic; antirheumatic; osteopathic;  
KW dermatological; antianaemic; antipsoriatic; hepatotropic; antigout;  
KW antinflammatory; antiHIV; protein transport regulator; cancer;  
KW immune disorder; cell proliferative disorder; secretory disorder;  
KW urticaria; allergy; abnormal vesicle trafficking; asthma;  
KW autoimmune haemolytic anaemia.  
XX Homo sapiens.  
OS  
XX WO200012703-A2.  
PN  
XX 09-MAR-2000.  
PD  
XX 26-AUG-1999; 99WO-US19616.  
PF  
XX 27-AUG-1998; 98US-0098206.  
PR  
XX (INCY-) INCYTE PHARM INC.  
PA  
XX Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;  
PI Gorgone GA, Baughn MR, Patterson C;

XX WPI; 2000-256642/22.  
 DR N-PSDB; AAA08040.  
 XX  
 XX New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis  
 XX  
 XX Claim 1; Page 67; 75pp; English.  
 PS  
 XX AAA08035 to AAA08042 encode the human protein transport-associated  
 CC molecules (PTAMs) given in AY82317 to AY82324. The PTAMs have  
 CC cytosolic, antiarthritic, antiasthmatic, immunosuppressant,  
 CC antitumor, antileukemic, antiallergic, antidiabetic, antilipemic,  
 CC antirheumatic, osteopathic, dermatological, antianemic, antipsoriatic,  
 CC hepatotropic, antitumor, antineoplastic and antihiv activities, and  
 CC regulate protein transport. PTAM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PTAM. PTAM polynucleotides are useful for  
 CC diagnosing conditions associated with PTAM, comprising detecting PTAM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the  
 CC biological sample. Diseases prevented, treated or diagnosed include cell  
 CC proliferative disorders such as cancers, immune disorders, secretory  
 CC disorders and other conditions associated with abnormal vesicle  
 CC trafficking, such as allergies, asthma, urticaria and autoimmune  
 CC haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists,  
 CC a targeting or delivery mechanism for bringing pharmaceutical agents  
 CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related  
 CC disorders. PTAM, its catalytic or immunogenic fragments are useful for  
 CC drug screening using libraries of compounds. PTAM polynucleotides are  
 CC useful for generating hybridisation probes useful in mapping the  
 CC naturally occurring genomic sequences.  
 XX  
 SQ Sequence 175 AA:  
 Query Match 100.0%; Score 936; DB 21; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-83;  
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGILLGALLLHGLTVDYGRPILEVPESVTGPKGVDNLPCTYDPLQGYTVLVKLVQR 60  
 DB 1 MGILLGALLLHGLTVDYGRPILEVPESVTGPKGVDNLPCTYDPLQGYTVLVKLVQR 60  
 QY 61 GSDPVTIFLRSSGDHIIQAKYQGRHSHKVPDVSLSLSTLEMDRSHYTCVETWQTP 120  
 DB 61 GSDPVTIFLRSSGDHIIQAKYQGRHSHKVPDVSLSLSTLEMDRSHYTCVETWQTP 120  
 QY 121 DGNQVVRDKITELRVQKHSSKLLTKTTEAPTMTYPLKATSTVKQSWDWTMDG 175  
 DB 121 DGNQVVRDKITELRVQKHSSKLLTKTTEAPTMTYPLKATSTVKQSWDWTMDG 175  
 RESULT 2  
 AAY30814  
 ID AAY30814 standard; Protein: 306 AA.  
 XX  
 XX AAY30814;  
 XX  
 XX 12-OCT-1999 (first entry)  
 XX  
 XX Human secreted protein encoded from gene 4.  
 DE  
 XX Secreted protein; prevention; treatment; protein therapy; gene therapy;  
 KW diagnosis; cancer; tumour; neurodegenerative disorder; blood disorder;  
 KW developmental abnormality; fetal deficiency; leukemia; autoimmune; acne;  
 KW hepatic disease; renal disease; lymphoma; inflammation; allergy; asthma;  
 KW Alzheimer's disease; cognitive disorder; schizophrenia; obesity; sepsis;  
 KW osteoporosis; arthritis; infection; AIDS; connective tissue disorder;  
 KW transplant rejection; diabetes; psoriasis; cardiovascular disorder;  
 KW reproductive disorder; food additive; food preservative; human; primer;  
 KW early promoter; GAS; gamma activation element.  
 XX  
 XX Homo sapiens.  
 OS

XX WO9940100-A1.  
 PN  
 XX  
 PD 12-AUG-1999.  
 XX  
 PF 04-FEB-1999; 99WO-US02293.  
 XX  
 PR 09-FEB-1998; 98US-0074341.  
 PR 09-FEB-1998; 98US-0074037.  
 PR 09-FEB-1998; 98US-0074118.  
 PR 09-FEB-1998; 98US-0074141.  
 PR 09-FEB-1998; 98US-0074157.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Kyaw H., Lafleur DW, Moore PA, Rosen CA, Ruben SM;  
 PI Shi Y, Wei Y;  
 DR WPI; 1999-479426/40.  
 DR N-PSDB; AAZ00805.  
 XX  
 PT New isolated human genes potentially useful for, e.g. developmental  
 PT abnormalities and fetal deficiencies  
 XX  
 XX Claim 1b; Page 209-210; 263pp; English.  
 XX  
 CC This invention describes novel isolated human genes and the secreted  
 CC proteins they encode. The polynucleotides and their corresponding  
 CC secreted polypeptides are useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. Also pathological  
 CC conditions can be diagnosed by determining the amount of the new  
 CC polypeptides in a sample or by determining the presence of mutations in  
 CC the new polynucleotides. Specific uses are described for the  
 CC polynucleotides of the invention based on which tissues they are most  
 CC highly expressed in, and include developing products for the diagnosis or  
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental  
 CC abnormalities and fetal deficiencies, blood diseases, leukemias,  
 CC diseases of the immune system, autoimmune diseases, hepatic and renal  
 CC disease, lymphomas, inflammation, allergies, Alzheimer's and cognitive  
 CC disorders, schizophrenia, obesity, osteoporosis, arthritis, infections,  
 CC AIDS, connective tissue disorders, transplant rejection, diabetes,  
 CC asthma, sepsis, acne, psoriasis, cardiovascular disorders, and  
 CC reproductive disorders. The polypeptides or polynucleotides can also be  
 CC used as food additives or preservatives. The polypeptide are also useful  
 CC for identifying their binding partners. This sequence represents a  
 CC secreted protein described in the invention.  
 XX  
 SQ Sequence 306 AA:  
 Query Match 100.0%; Score 936; DB 20; Length 306;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-82;  
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGILLGALLLHGLTVDYGRPILEVPESVTGPKGVDNLPCTYDPLQGYTVLVKLVQR 60  
 DB 1 MGILLGALLLHGLTVDYGRPILEVPESVTGPKGVDNLPCTYDPLQGYTVLVKLVQR 60  
 QY 61 GSDPVTIFLRSSGDHIIQAKYQGRHSHKVPDVSLSLSTLEMDRSHYTCVETWQTP 120  
 DB 61 GSDPVTIFLRSSGDHIIQAKYQGRHSHKVPDVSLSLSTLEMDRSHYTCVETWQTP 120  
 QY 121 DGNQVVRDKITELRVQKHSSKLLTKTTEAPTMTYPLKATSTVKQSWDWTMDG 175  
 DB 121 DGNQVVRDKITELRVQKHSSKLLTKTTEAPTMTYPLKATSTVKQSWDWTMDG 175  
 RESULT 3  
 AAY41691  
 ID AAY41691 standard; Protein: 321 AA.  
 XX  
 XX AAY41691;  
 XX  
 XX 07-DEC-1999 (first entry)  
 DT



XX DE Human PRO 362 protein sequence.  
 XX KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
 KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
 XX secreted protein; transmembrane protein.  
 OS Homo sapiens.  
 XX PN WO9946281-A2.  
 XX PD 16-SEP-1999.  
 XX PF 08-MAR-1999; 99WO-US05028.  
 XX 10-MAR-1998; 98US-0077450.  
 PR 11-MAR-1998; 98US-0077632.  
 PR 11-MAR-1998; 98US-0077641.  
 PR 11-MAR-1998; 98US-0077649.  
 PR 12-MAR-1998; 98US-0077791.  
 PR 13-MAR-1998; 98US-0078004.  
 PR 17-MAR-1998; 98US-0040220.  
 PR 20-MAR-1998; 98US-0078886.  
 PR 20-MAR-1998; 98US-0078910.  
 PR 20-MAR-1998; 98US-0078936.  
 PR 20-MAR-1998; 98US-0078939.  
 PR 25-MAR-1998; 98US-0079294.  
 PR 26-MAR-1998; 98US-0079656.  
 PR 27-MAR-1998; 98US-0079664.  
 PR 27-MAR-1998; 98US-0079689.  
 PR 27-MAR-1998; 98US-0079728.  
 PR 27-MAR-1998; 98US-0079786.  
 PR 30-MAR-1998; 98US-0079920.  
 PR 30-MAR-1998; 98US-0079923.  
 PR 31-MAR-1998; 98US-0080105.  
 PR 31-MAR-1998; 98US-0080107.  
 PR 31-MAR-1998; 98US-0080165.  
 PR 31-MAR-1998; 98US-0080194.  
 PR 01-APR-1998; 98US-0080327.  
 PR 01-APR-1998; 98US-0080328.  
 PR 01-APR-1998; 98US-0080333.  
 PR 01-APR-1998; 98US-0080334.  
 PR 01-APR-1998; 98US-0081049.  
 PR 08-APR-1998; 98US-0081070.  
 PR 08-APR-1998; 98US-0081071.  
 PR 09-APR-1998; 98US-0081195.  
 PR 09-APR-1998; 98US-0081203.  
 PR 09-APR-1998; 98US-0081229.  
 PR 15-APR-1998; 98US-0081817.  
 PR 15-APR-1998; 98US-0081838.  
 PR 15-APR-1998; 98US-0081952.  
 PR 15-APR-1998; 98US-0081955.  
 PR 21-APR-1998; 98US-0082568.  
 PR 21-APR-1998; 98US-0082569.  
 PR 22-APR-1998; 98US-0082700.  
 PR 22-APR-1998; 98US-0082704.  
 PR 23-APR-1998; 98US-0082767.  
 PR 23-APR-1998; 98US-0082804.  
 PR 23-APR-1998; 98US-0082796.  
 PR 27-APR-1998; 98US-0083336.  
 PR 28-APR-1998; 98US-0083322.  
 PR 29-APR-1998; 98US-0083392.  
 PR 29-APR-1998; 98US-0083495.  
 PR 29-APR-1998; 98US-0083496.  
 PR 29-APR-1998; 98US-0083499.  
 PR 29-APR-1998; 98US-0083500.  
 PR 29-APR-1998; 98US-0083545.  
 PR 29-APR-1998; 98US-0083554.  
 PR 29-APR-1998; 98US-0083558.  
 PR 29-APR-1998; 98US-0083559.  
 PR 30-APR-1998; 98US-0083742.  
 PR 05-MAY-1998; 98US-0084366.

PR 06-MAY-1998; 98US-0084414.  
 PR 06-MAY-1998; 98US-0084441.  
 PR 07-MAY-1998; 98US-0084598.  
 PR 07-MAY-1998; 98US-0084600.  
 PR 07-MAY-1998; 98US-0084627.  
 PR 07-MAY-1998; 98US-0084637.  
 PR 07-MAY-1998; 98US-0084639.  
 PR 07-MAY-1998; 98US-0084640.  
 PR 07-MAY-1998; 98US-0084643.  
 PR 13-MAY-1998; 98US-0085323.  
 PR 13-MAY-1998; 98US-0085338.  
 PR 13-MAY-1998; 98US-0085339.  
 PR 15-MAY-1998; 98US-0085573.  
 PR 15-MAY-1998; 98US-0085579.  
 PR 15-MAY-1998; 98US-0085580.  
 PR 15-MAY-1998; 98US-0085582.  
 PR 15-MAY-1998; 98US-0085689.  
 PR 15-MAY-1998; 98US-0085697.  
 PR 15-MAY-1998; 98US-0085700.  
 PR 15-MAY-1998; 98US-0085704.  
 PR 18-MAY-1998; 98US-0086023.  
 PR 22-MAY-1998; 98US-0086392.  
 PR 22-MAY-1998; 98US-0086414.  
 PR 22-MAY-1998; 98US-0086430.  
 PR 22-MAY-1998; 98US-0086486.  
 PR 28-MAY-1998; 98US-0087098.  
 PR 28-MAY-1998; 98US-0087106.  
 PR 28-MAY-1998; 98US-0087208.  
 PR 30-JUL-1998; 98US-0094651.  
 PR 11-SEP-1998; 98US-0100038.  
 XX (GETH ) GENENTECH INC.  
 PA Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;  
 XX WPI: 1999-551358/46.  
 DR N-PSDB; AAZ33935.  
 XX  
 PT New secreted and transmembrane polypeptides and their polynucleotides,  
 PT useful for treating blood coagulation disorders, cancers and cellular  
 PT adhesion disorders  
 XX  
 PS Claim 12; Fig 22; 530pp; English.  
 CC The present invention describes secreted and transmembrane polypeptides  
 CC and their polynucleotides. The nucleotide sequences are useful as  
 CC sources of probes, primers, for chromosome mapping, and for generation  
 CC of antisense sequences. They can also be used to create transgenic  
 CC animals. The proteins can be used to treat a variety of diseases and  
 CC disorders, depending on their function. Diseases that may be treated  
 CC include blood coagulation disorders, cancers and cellular adhesion  
 CC disorders. They may also be used to raise antibodies. AAZ33891 to  
 CC AAZ34338, and AA41685 to AA41774 represent polynucleotide and  
 CC polypeptide sequence given in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 321 AA;  
 Query Match 92.9%; Score 870; DB 20; Length 321;  
 Best Local Similarity 64.7%; Pred. No. 3.7e-76;  
 Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;  
 QY 1 MGILLGLLLGLHLYTDYGRPILEVPESTGPKGDNLPCTYDPLQGYTVLVKWLVR 60  
 DB 1 MGILLGLLLGLHLYTDYGRPILEVPESTGPKGDNLPCTYDPLQGYTVLVKWLVR 60  
 QY 61 GSDPVTIFLRDSSGGHIOQAQYGRHLVSHKVPDGVSLSTLEMDRSHYTCEVTWQPP 120  
 DB 61 GSDPVTIFLRDSSGGHIOQAQYGRHLVSHKVPDGVSLSTLEMDRSHYTCEVTWQPP 120  
 QY 121 DGNQVVRDKITELRVQ----- 136  
 DB 121 DGNQVVRDKITELRVQ----- 136

QY 137 -----KHSKLLKTK 146  
 DB 181 QTNQEPKIVATLTLFLKPAVIADSGSYFCTAKQGVSEHQSHDIVKFWKDSKLLKTK 240  
 QY 147 TEAPTMTYPLKATSTVKQSWDWTMDG 175  
 DB 241 TEAPTMTYPLKATSTVKQSWDWTMDG 269

RESULT 4  
 AAY23322  
 ID AAY23322 standard; protein: 321 AA.  
 XX  
 AC AAY23322;  
 XX  
 DT 02-SEP-1999 (first entry)  
 XX  
 DE Amino acid sequence of the PRO362 polypeptide.  
 XX  
 AA33 related antigen; PRO301; PRO362; PRO245; inflammatory disease;  
 KW tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09927098-A2.  
 PD  
 XX  
 PF 03-JUN-1999.  
 XX  
 XX 20-NOV-1998; 98WO-US24855.  
 XX  
 PR 17-SEP-1998; 98WO-US19437.  
 PR 21-NOV-1997; 97US-0066364.  
 PR 20-MAR-1998; 98US-0078936.  
 XX  
 XX (GETH ) GENENTECH INC.  
 PA  
 XX Ashkenazi A, Fong S, Goddard A, Gurney AL, Napier MA;  
 PI Tumas D, Wood WI;  
 XX  
 DR WPI: 1999-404743/34.  
 XX  
 XX Antigen PRO301, PRO362 and PRO245 related to A33  
 PT  
 XX  
 PS Claim 47; Fig 3; 122pp; English.  
 XX  
 CC The specification describes A33 related antigens PRO301, PRO362 and  
 CC PRO245. The methods and compositions of the invention are useful for the  
 CC treatment and diagnosis of inflammatory disease and tumours in mammals.  
 CC Such inflammatory diseases include of inflammatory bowel disease,  
 CC systemic lupus erythematosus, rheumatoid arthritis, juvenile chronic  
 CC arthritis, spondyloarthropathies, systemic sclerosis, scleroderma,  
 CC idiopathic inflammatory myopathies, dermatomyositis, polymyositis,  
 CC Sjogren's syndrome, systemic vaculitis, sarcoidosis, autoimmune hemolytic  
 CC anemia, immune pancytopenia, paroxysmal nocturnal hemoglobinuria,  
 CC autoimmune thrombocytopenia, idiopathic thrombocytopenic purpura,  
 CC immune-mediated thrombocytopenia, thyroiditis, Grave's disease,  
 CC Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic  
 CC thyroiditis, diabetes mellitus, immune-mediated renal disease,  
 CC glomerulonephritis, tubulointerstitial nephritis, demyelinating diseases  
 CC of the central and peripheral nervous systems such as multiple sclerosis,  
 CC idiopathic polyneuropathy, hepatobiliary diseases, infectious hepatitis  
 CC A, B, C, D, E, nonhepatotropic viruses, autoimmune chronic active  
 CC hepatitis, primary biliary cirrhosis, granulomatous hepatitis, sclerosing  
 CC cholangitis, inflammatory and fibrotic lung diseases, gluten-sensitive  
 CC enteropathy, Whipple's disease, autoimmune or immune-mediated skin  
 CC diseases allergic diseases of the lung such as eosinophilic pneumonias,  
 CC idiopathic pulmonary fibrosis and hypersensitivity pneumonitis  
 CC CC transplantation associated diseases disease. The present sequence  
 CC represents the PRO362 polypeptide.  
 XX  
 XX Sequence 321 AA;

Query Match 92.9%; Score 870; DB 20; Length 321.  
 Best Local Similarity 64.7%; Pred. No. 3.7e-76;  
 Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;  
 QY 1 MGILLGLLGLHLAVDTYGRPILEVPEVPGWGDVNLCTYDPLQGYTQVLVKWLVR 60  
 DB 1 MGILLGLLGLHLAVDTYGRPILEVPEVPGWGDVNLCTYDPLQGYTQVLVKWLVR 60  
 QY 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSLSLSTLEMDDRSHYTCVETWQTP 120  
 DB 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSLSLSTLEMDDRSHYTCVETWQTP 120  
 QY 121 DGNQVVRDKITELRVO----- 136  
 DB 121 DGNQVVRDKITELRVQKLSVSKPTVTGSGYFTVPGMRISLQOQARGSPISYIWKQ 180  
 QY 137 -----KHSSKLLKTK 146  
 DB 181 QTNQEPKIVATLTLFLKPAVIADSGSYFCTAKQGVSEHQSHDIVKFWKDSKLLKTK 240  
 QY 147 TEAPTMTYPLKATSTVKQSWDWTMDG 175  
 DB 241 TEAPTMTYPLKATSTVKQSWDWTMDG 269

RESULT 5  
 AAB44247  
 ID AAB44247 standard; Protein: 321 AA.  
 XX  
 AC AAB44247;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human PRO362 (UNQ317) protein sequence SEQ ID NO:52.  
 XX  
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;  
 XX expressed sequence tag; detection: cancer.  
 OS Homo sapiens.  
 XX  
 PN W0200053756-A2.  
 PD  
 XX  
 PF 14-SEP-2000.  
 XX  
 PF 18-FEB-2000; 2000WO-US04341.  
 XX  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 29-MAR-1999; 99US-0126773.  
 PR 21-APR-1999; 99US-0130232.  
 PR 28-APR-1999; 99US-0131445.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 16-DEC-1999; 99WO-US28565.  
 PR 30-DEC-1999; 99WO-US30093.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 05-JAN-2000; 99WO-US31274.  
 PR 06-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 XX  
 XX (GETH ) GENENTECH INC.  
 PA  
 XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;  
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
 PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;  
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;  
 XX

DR WPI; 2000-611443/58.  
DR N-PSDB; AAC78472.  
XX  
PT Novel PRO polypeptides and polynucleotides used in detection methods,  
PT to target bioactive molecules to specific cells, and to modulate  
PT cellular activities  
XX  
XX Claim 12; Fig 22; 636pp; English.  
XX  
CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed  
CC sequence tag) sequences which encode secreted or transmembrane PRO  
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
CC activity. The polynucleotides and polypeptides can be used for detecting  
CC the presence of PRO polypeptides in samples, for linking bioactive  
CC molecules to cells and for modulating biological activities of cells,  
CC using the polypeptides for specific targeting. The polypeptide targeting  
CC can be used to kill the target cells, e.g. for the treatment of cancers.  
CC The polypeptide pairs provide specific targeting of bioactive molecules  
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in  
CC the isolation of the PRO polynucleotide sequences.  
XX  
SQ Sequence 321 AA;  
Query Match 92.9%; Score 870; DB 21; Length 321;  
Best Local Similarity 64.7%; Pred. No. 3.7e-76;  
Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;  
QY 1 MGILLGILLGHLTVDTYGRPILEVPSVTGPKGDVNLPCCTYDPLQGYTQVLVQLVQR 60  
DB 1 MGILLGILLGHLTVDTYGRPILEVPSVTGPKGDVNLPCCTYDPLQGYTQVLVQLVQR 60  
QY 61 GSDPVTIFLRDSSGDHIQQAQYGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTQTP 120  
DB 61 GSDPVTIFLRDSSGDHIQQAQYGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTQTP 120  
QY 121 DGNOVVRDKITELRVQ----- 136  
DB 121 DGNOVVRDKITELRVQKLSVSKPTVTGSGYFTVPGMRSISLQCGARGSPPISYIWKQ 180  
QY 137 -----KHSSKLLKTK 146  
DB 137 -----KHSSKLLKTK 146  
DB 181 QTNQEPKIVATLSTLLFKPAVIADSGSYFCTAKGVGSGQHSQSDIVKVFVYKDSKLLKTK 240  
QY 147 TEAPTTMTYPLKATSTVKQSWDWTMDG 175  
DB 241 TEAPTTMTYPLKATSTVKQSWDWTMDG 269  
RESULT 6  
AAB24047  
ID AAB24047 standard; Protein; 321 AA.  
XX  
AC AAB24047;  
XX  
DT 25-JAN-2001 (first entry)  
XX  
DE Human PRO362 protein sequence SEQ ID NO:14.  
XX  
KW Human; tumour; diagnosis; neoplastic disease; identification; cancer;  
KW tumorigenesis; detection; neoplastic cell growth; proliferation;  
KW cytostatic; antiinflammatory; immunomodulatory; inflammatory disorder;  
KW immunological disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200053754-A1.  
XX  
PD 14-SEP-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00277.  
XX  
PR 08-MAR-1999; 99WO-US05028.  
PR 12-MAR-1999; 99US-0123957.

PR 29-MAR-1999; 99US-0126773.  
PR 21-APR-1999; 99US-0130232.  
PR 28-APR-1999; 99US-0131445.  
PR 05-OCT-1999; 99WO-US23089.  
PR 30-NOV-1999; 99WO-US28313.  
PR 02-DEC-1999; 99WO-US28551.  
PR 02-DEC-1999; 99WO-US28564.  
PR 30-DEC-1999; 99WO-US31243.  
PR 30-DEC-1999; 99WO-US31274.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Baker KP, Desauvage FJ, Goddard A, Gurney AL, Klein RD, Roy MA;  
PI Wood WI;  
XX  
DR WPI; 2000-572269/53.  
DR N-PSDB; AAC58231.  
XX  
PT New isolated antibody for use in compositions and methods for the  
PT diagnosis and treatment of neoplastic cell growth and proliferation in  
PT mammals, including humans, and in monitoring tumor treatment  
XX  
PS Claim 61; Fig 14; 195pp; English.  
XX  
CC The present invention describes an isolated antibody (Ab) that binds to  
CC one of the human proteins (P) designated PRO1330, PRO1449,  
CC PRO237, PRO324, PRO351, PRO362, PRO615, PRO538, PRO3664, PRO618,  
CC PRO772, PRO703, PRO792 or PRO474. The Ab can be used in compositions  
CC and methods for the diagnosis and treatment of neoplastic cell growth  
CC and proliferation in mammals, including humans. Genes and polypeptides  
CC encoded by them, that are amplified in the genome of a tumour cell, can  
CC be identified and are useful targets for the treatment and prevention of  
CC certain cancers and may be used to monitor tumour treatment. Compounds  
CC that inhibit the expression or activity of the identified polypeptides  
CC can be identified and used as antagonists. Benign or malignant tumours,  
CC inflammatory disorders and immunological disorders can be treated.  
CC AAC58123 to AAC58224 represent hybridisation probes and PCR primers used  
CC in the isolation of the human PRO sequences. AAC58225 to AAC58241 and  
CC AAB24041 to AAB24056 represent human PRO polynucleotide and protein  
CC sequences given in the exemplification of the present invention.  
XX  
SQ Sequence 321 AA;  
Query Match 92.9%; Score 870; DB 21; Length 321;  
Best Local Similarity 64.7%; Pred. No. 3.7e-76;  
Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;  
QY 1 MGILLGILLGHLTVDTYGRPILEVPSVTGPKGDVNLPCCTYDPLQGYTQVLVQLVQR 60  
DB 1 MGILLGILLGHLTVDTYGRPILEVPSVTGPKGDVNLPCCTYDPLQGYTQVLVQLVQR 60  
QY 61 GSDPVTIFLRDSSGDHIQQAQYGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTQTP 120  
DB 61 GSDPVTIFLRDSSGDHIQQAQYGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTQTP 120  
QY 121 DGNOVVRDKITELRVQ----- 136  
DB 121 DGNOVVRDKITELRVQKLSVSKPTVTGSGYFTVPGMRSISLQCGARGSPPISYIWKQ 180  
QY 137 -----KHSSKLLKTK 146  
DB 181 QTNQEPKIVATLSTLLFKPAVIADSGSYFCTAKGVGSGQHSQSDIVKVFVYKDSKLLKTK 240  
QY 147 TEAPTTMTYPLKATSTVKQSWDWTMDG 175  
DB 241 TEAPTTMTYPLKATSTVKQSWDWTMDG 269  
RESULT 7  
AAB33429  
ID AAB33429 standard; Protein; 321 AA.  
XX  
AC AAB33429;

XX DT 29-JAN-2001 (first entry)

XX DE Human PRO362 protein UNQ317 SEQ ID NO:80.

XX KW Human; immune related disease; diagnosis; antiinflammatory; cardiant;

KW dermatological; antiarthritic; antirheumatic; immunosuppressive;

KW haemostatic; antithyroid; antidiabetic; nootropic; neuroprotective;

KW antianemic; hepatotropic; virucide; antipsoriatic; anti allergic;

KW antiasthmatic; systemic lupus erythematosus; rheumatoid arthritis;

KW osteoarthritis; spondyloarthropathy; systemic sclerosis; sarcoidosis;

KW idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;

KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;

KW autoimmune thrombocytopenia; immune-mediated renal disease;

KW demyelinating disease; hepatobiliary disease; Whipple's disease;

KW inflammatory bowel disease; gluten-sensitive enteropathy;

KW autoimmune disease; immune-mediated skin disease; allergic disease;

KW immunological disease; transplantation associated disease;

KW graft rejection; graft-versus-host-disease.

XX OS Homo sapiens.

XX KW WO200053758-A2.

XX PN 14-SEP-2000.

XX PD 02-MAR-2000; 2000WO-US05841.

XX PE 08-MAR-1999; 99WO-US05028.

XX PR 10-MAR-1999; 99US-0123618.

PR 12-MAR-1999; 99US-0123957.

PR 23-MAR-1999; 99US-0125775.

PR 12-APR-1999; 99US-0128849.

PR 20-APR-1999; 99WO-US08615.

PR 28-APR-1999; 99US-0131445.

PR 04-MAY-1999; 99US-0132371.

PR 14-MAY-1999; 99US-0134287.

PR 02-JUN-1999; 99WO-US12252.

PR 23-JUN-1999; 99US-0141037.

PR 20-JUL-1999; 99US-0144758.

PR 26-JUL-1999; 99US-0145698.

PR 28-JUL-1999; 99US-0146222.

PR 01-SEP-1999; 99WO-US20111.

PR 08-SEP-1999; 99WO-US20594.

PR 13-SEP-1999; 99WO-US20944.

PR 15-SEP-1999; 99WO-US21090.

PR 15-SEP-1999; 99WO-US21547.

PR 05-OCT-1999; 99WO-US23089.

PR 29-OCT-1999; 99US-0162506.

PR 29-NOV-1999; 99WO-US28214.

PR 30-NOV-1999; 99WO-US28313.

PR 30-NOV-1999; 99WO-US28409.

PR 01-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US28634.

PR 02-DEC-1999; 99WO-US28551.

PR 02-DEC-1999; 99WO-US28564.

PR 16-DEC-1999; 99WO-US28565.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30999.

PR 30-DEC-1999; 99WO-US31274.

PR 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00277.

PR 06-JAN-2000; 2000WO-US00376.

PR 11-FEB-2000; 2000WO-US03565.

PR 18-FEB-2000; 2000WO-US04341.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

XX (GETH ) GENENTECH, INC.

XX AShenkazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W;

PI Kabakoff RC, Lu Y, Pan J, Pennica D, Shelton DL, Smith V;

PI Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M;

XX WPI; 2000-572271/53.

DR N-PSDB; AAC58594.

XX PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of

PT immune related disorders, e.g. systemic lupus erythematosus, rheumatoid

PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -

PS Claim 33; Fig 32; 309pp; English.

XX CC The present invention describes sixty four human PRO proteins which can

CC be used in the treatment of immune related diseases. The human PRO

CC proteins, anti-PRO antibodies, agonists and antagonists are useful for

CC treating and diagnosing immune related disorders. The disorders are

CC selected from systemic lupus erythematosus, rheumatoid arthritis,

CC osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,

CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's

CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic

CC anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus,

CC immune-mediated renal disease, demyelinating diseases of the central

CC and peripheral nervous systems, hepatobiliary diseases, inflammatory

CC bowel disease, gluten-sensitive enteropathy and Whipple's disease,

CC autoimmune or immune-mediated skin diseases, allergic diseases,

CC immunological diseases of the lung, and transplantation associated

CC diseases including graft rejection and graft-versus-host-disease.

CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used

CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and

CC AAB33414 to AAB33477 represent human PRO polynucleotide and protein

CC sequences given in the exemplification of the present invention.

XX SQ Sequence 321 AA;

Query Match 92.9%; Score 870; DB 21; Length 321;

Best Local Similarity 64.7%; Pred. No. 3.7e-76;

Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

QY 1 MGILLGLLLGLLHGLTVDYGRPILEVPEVSGVPGWGDVNLPCYDPLQGYTVLVKWLVR 60

DB 1 MGILLGLLLGLLHGLTVDYGRPILEVPEVSGVPGWGDVNLPCYDPLQGYTVLVKWLVR 60

QY 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVNTWQP 120

DB 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVNTWQP 120

QY 121 DGNQVVRDKITELRVQ----- 136

DB 121 DGNQVVRDKITELRVQKLSVSKPTVTGSGYGTVPQGMRLSLQCOARGSPPISTIWKQ 180

QY 137 -----KHSSKLLKTK 146

DB 181 QTNNOEPIKVAITLSLLFKPAVIADSGSYFCTAKGVGSEQHSDIVKVFVKDSSKLLKTK 240

QY 147 TEAPTMTYPLKATSTVKOSWDWTTDMG 175

DB 241 TEAPTMTYPLKATSTVKOSWDWTTDMG 269

RESULT 8

AA95346

ID AA95346 standard; Protein; 321 AA.

XX AC AA95346;

XX DT 25-SEP-2000 (first entry)

XX DE Human PRO362 antitumour protein.

XX KW PRO362; human; antitumour; tumour; therapy; cytostatic;

KW breast cancer; ovarian cancer; renal cancer; colorectal cancer;

KW uterine cancer; prostate cancer; lung cancer; bladder cancer;

XX central nervous system cancer; melanoma; leukaemia; neoplasm.

OS Homo sapiens.

XX PH Key Location/Qualifiers  
 FT Peptide 1..19  
 FT Protein /label= Signal\_peptide  
 FT Protein 20..321  
 FT Domain /label= PRO362  
 FT Domain 281..300  
 FT Region /note= "transmembrane domain"  
 FT Region 149..153  
 FT Modified-site /note= "glycosaminoglycan attachment site"  
 FT Modified-site 308..312  
 FT Modified-site /note= "protein kinase phosphorylation site"  
 FT Modified-site 2..8  
 FT Modified-site /note= "N-myristoylated"  
 FT Modified-site 148..154  
 FT Modified-site /note= "N-myristoylated"  
 FT Modified-site 158..164  
 FT Modified-site /note= "N-myristoylated"  
 FT Modified-site 207..213  
 FT Modified-site /note= "N-myristoylated"  
 FT Modified-site 215..221  
 FT Modified-site /note= "N-myristoylated"  
 XX WO200037638-A2.  
 PN 29-JUN-2000.  
 XX 02-DEC-1999; 99WO-US28555.  
 XX 22-DEC-1998; 98US-0113296.  
 XX 08-MAR-1999; 99WO-US05028.  
 XX 21-APR-1999; 99US-0130232.  
 XX 28-APR-1999; 99US-0131445.  
 XX 14-MAY-1999; 99US-0134287.  
 XX 20-JUL-1999; 99US-0144758.  
 XX 26-JUL-1999; 99US-0145698.  
 XX 15-SEP-1999; 99WO-US21090.  
 XX 15-SEP-1999; 99WO-US21547.  
 XX (GETH ) GENENTECH INC.  
 XX Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Marsters SA;  
 PI Napier MA, Pitti RM, Wood WI;  
 XX WPI: 2000-442668/38..  
 DR N-PSDB; AAA49725.  
 XX Novel composition to inhibit neoplastic cell growth or for treating  
 PT tumor in mammal comprises polypeptides PRO179, PRO207, PRO320, PRO219,  
 PT PRO221, PRO224, PRO328, PRO301, PRO526, PRO362, PRO356, PRO509 or  
 PT PRO866  
 XX Claim 19; Fig 20; 172pp; English.  
 XX The present sequence is that of human antitumor protein PRO362, as  
 CC deduced from a foetal brain cDNA clone (see AAA49725). PRO362 has a  
 CC mol.wt. of 35,544 and a pI of 8.51. It has homology to A33 antigen  
 CC protein and to the HCAR protein as well as to Dayoff sequences. A  
 CC claimed method for inhibiting the growth of a tumour cell comprises  
 CC exposing the tumor cell to PRO179, PRO207, PRO320, PRO219, PRO221  
 CC PRO224, PRO328, PRO301, PRO356, PRO362, PRO509 or PRO866 (see  
 CC AAY5337-49); their agonists or chimeric polypeptides incorporating  
 CC them. The tumour is especially a cancer selected from breast, ovarian,  
 CC renal, colorectal, uterine, prostate, lung, bladder and central nervous  
 CC system cancer, melanoma and leukaemia. Methods for the recombinant  
 CC expression of the antitumour proteins are also provided.  
 XX Sequence 321 AA;  
 SQ Query Match 92.98; Score 870; DB 21; Length 321;  
 Best Local Similarity 64.78; Pred. No. 3.7e-76;  
 Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

QY 1 MGILLGILLGHLTVDTYGRPILEVPESVTGPWKGDNLPCTYDPLOGYTOVLVKWLVR 60  
 DB 1 MGILLGILLGHLTVDTYGRPILEVPESVTGPWKGDNLPCTYDPLOGYTOVLVKWLVR 60  
 QY 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSLSQLSTLEMDRSHYTCVWTQTP 120  
 DB 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSLSQLSTLEMDRSHYTCVWTQTP 120  
 QY 121 DGNOVVRDKITELRVQ----- 136  
 DB 121 DGNOVVRDKITELRVQKLSVSKPTVTGSGYGTVPQGMRIQLQCARQSPPIYINVKQ 180  
 QY 137 -----KHSSKLLKTK 146  
 DB 181 QTNQDEPIKVATLSTLLFKPAVIADSGSYFCTAKGVQSEQHSDIVFVKDSSKLLKTK 240  
 QY 147 TEAPTTMTYPLKATSTVKQSWDWTMDG 175  
 DB 241 TEAPTTMTYPLKATSTVKQSWDWTMDG 269  
 RESULT 9  
 AAB19396  
 ID AAB19396 standard; Protein; 399 AA.  
 XX AC AAB19396;  
 XX DT 06-MAR-2001 (first entry)  
 XX DE Amino acid sequence of a human secreted protein.  
 XX KW Secreted protein; platelet disorder; stem cell disorder; osteoporosis;  
 KW osteoarthritis; burn; incision; ulcer; periodontal disease; neuropathy;  
 KW nervous system disease; bone growth; cosmetic plastic surgery;  
 KW gut protection; gut regeneration; fibrosis; cancer;  
 KW bone marrow transplantation.  
 XX OS Homo sapiens.  
 XX WO2000061755-A2.  
 XX PD 19-OCT-2000.  
 XX PF 10-APR-2000; 2000WO-US09555.  
 XX PR 09-APR-1999; 99US-0128574.  
 XX PR 20-AUG-1999; 99US-0150054.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PI Garcia PD;  
 XX WPI: 2000-665133/64.  
 XX DR N-PSDB; AAC61882.  
 XX Novel secreted human proteins useful for stimulating blood cell  
 PT generation in patients receiving cancer chemotherapy, treating bone  
 PT marrow transplantation patients and for healing fractured bones  
 XX Claim 14; Page 57; 74pp; English.  
 XX AAB19393-B19407 represent secreted human proteins. The secreted proteins  
 CC are useful in assays to determine their biological activities. The  
 CC proteins can also be used as biomarkers to identify tissues or cell  
 CC types which express the proteins. The polynucleotide molecules can be  
 CC used as biomarkers for tissues or chromosomes and to elicit immune  
 CC responses. The proteins and antibodies are useful in diagnosis and  
 CC treatment of diseases associated with altered expression of these  
 CC proteins. The proteins are also useful for prevention or treatment of  
 CC platelet disorders, stem cell disorders, osteoporosis or osteoarthritis,  
 CC burns, incisions, ulcers, periodontal diseases, central and peripheral  
 CC nervous system diseases and neuropathies, for healing fractured bones  
 CC and to induce cartilage and/or bone growth in cosmetic plastic surgery.

CC The proteins are also useful for gut protection or regeneration, for the  
 CC treatment of lung or liver fibrosis, for stimulating blood cell  
 CC generation in patients receiving cancer chemotherapy and for treatment  
 CC of bone marrow transplantation patients.

XX Sequence 399 AA;

SQ Query Match 92.9%; Score 870; DB 21; Length 399;  
 Best Local Similarity 64.7%; Pred. No. 5e-76;  
 Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

QY 1 MGILLGILLGLHGLTVDYGRPILEVPSVTPGPKGDVNLPCYDPLQGYQVLYKWLVR 60  
 DB 1 MGILLGILLGLHGLTVDYGRPILEVPSVTPGPKGDVNLPCYDPLQGYQVLYKWLVR 60  
 QY 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTQTP 120  
 DB 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTQTP 120  
 QY 121 DGNQVVRDKITELRVQ----- 136  
 DB 121 DGNQVVRDKITELRVQ----- 136  
 QY 121 QTNQEPKIVATLSTLFLKPAVIADSGYFCTAKGVQSGHSDIVFVVKDSSKLLTK 146  
 DB 121 QTNQEPKIVATLSTLFLKPAVIADSGYFCTAKGVQSGHSDIVFVVKDSSKLLTK 146  
 QY 147 TEAPTTMTYPLKATSTVKQSWDWTMDG/175  
 DB 241 TEAPTTMTYPLKATSTVKQSWDWTMDG 269

# RESULT 10

AAM93874  
 ID AAM93874 standard; Protein; 399 AA.

XX AAM93874;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 3986.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

XX N-PSDB; AAK94833.

PT 830 Primers useful for synthesizing full length cDNA clones and their  
 PT use in genetic manipulation

XX Claim 8; SEQ ID NO 3986; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA  
 CC clones. 830 cDNA molecules encoding a human protein have been  
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
 CC molecules have been determined. Primers for synthesizing the full length

CC cDNA are useful for clarifying the function of the protein encoded by  
 CC the cDNA. The full length clones were obtained by construction of full  
 CC length enriched cDNA libraries that were synthesised by the oligo-capping  
 CC method. The primers enable the production of the full length cDNA easily  
 CC without any special methods. The present sequence is a polypeptide  
 CC encoded by a full length human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 399 AA;

Query Match 92.9%; Score 870; DB 22; Length 399;  
 Best Local Similarity 64.7%; Pred. No. 5e-76;  
 Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

QY 1 MGILLGILLGLHGLTVDYGRPILEVPSVTPGPKGDVNLPCYDPLQGYQVLYKWLVR 60  
 DB 1 MGILLGILLGLHGLTVDYGRPILEVPSVTPGPKGDVNLPCYDPLQGYQVLYKWLVR 60  
 QY 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTQTP 120  
 DB 61 GSDPVTIFLRDSSGDHIQQAQYQGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTQTP 120  
 QY 121 DGNQVVRDKITELRVQ----- 136  
 DB 121 DGNQVVRDKITELRVQ----- 136  
 QY 137 -----KHSSKLLTK 146  
 DB 181 QTNQEPKIVATLSTLFLKPAVIADSGYFCTAKGVQSGHSDIVFVVKDSSKLLTK 240  
 QY 147 TEAPTTMTYPLKATSTVKQSWDWTMDG 175  
 DB 241 TEAPTTMTYPLKATSTVKQSWDWTMDG 269

# RESULT 11

AAE04208

ID AAE04208 standard; Protein; 399 AA.

XX AAE04208;

DT 09-AUG-2001 (first entry)

XX Human gene 11 encoded secreted protein HMSOW51, SEQ ID NO:63.

XX Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 KW endocrine disorder; infection; wound healing; vulnery;  
 KW cell culture; chemotaxis; food additive; chromosome X;  
 KW binding partner identification.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..19

XX Protein 20..399

XX /note= "Mature secreted protein"

XX WO200136432-A2.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000WO-US31162.

XX 19-NOV-1999; 99US-0166415.

```

PR 30-JUN-2000; 2000US-0215136.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;
XX
DR WPI; 2001-343793/36.
DR N-PSDB; AAD08498.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition
PT
XX
XX Claim 11; Page 440-441; 509pp; English.
XX
XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.
CC AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 18 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.
XX
XX Sequence 399 AA;
SQ
Query Match. 92.9%; Score 870; DB 22; Length 399;
Best Local Similarity 64.7%; Pred. No. 5e-76;
Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;
QY 1 MGILLGLLLGLHVTDTYGRPILEVPESTGPKGVDNLPCTYDPLQGYTQVLKVLVQR 60
DB 1 MGILLGLLLGLHVTDTYGRPILEVPESTGPKGVDNLPCTYDPLQGYTQVLKVLVQR 60
QY 61 GSDPVTIFLRDSSGDHIQAKYQGRHVSHPKPGDVSLSLSTLEMDRSHYTCVETWQTP 120
DB 61 GSDPVTIFLRDSSGDHIQAKYQGRHVSHPKPGDVSLSLSTLEMDRSHYTCVETWQTP 120
QY 121 DGNQVVRDKITELRVO----- 136
DB 121 DGNQVVRDKITELRVO----- 136
QY 121 DGNQVVRDKITELRVO-----KHSSKLLKTK 146
DB 121 DGNQVVRDKITELRVO-----KHSSKLLKTK 146
QY 137 -----KHSSKLLKTK 146
DB 137 -----KHSSKLLKTK 146
QY 147 TEAPTTWTYPLKATSVKQSWDWTMDMG 175
DB 147 TEAPTTWTYPLKATSVKQSWDWTMDMG 175
QY 241 TEAPTTWTYPLKATSVKQSWDWTMDMG 269
DB 241 TEAPTTWTYPLKATSVKQSWDWTMDMG 269
RESULT 12
AAE04230

```

```

ID AAE04230 standard; Protein; 399 AA.
XX
XX AAE04230;
XX
XX 09-AUG-2001 (first entry)
XX
XX Human gene 11 encoded secreted protein HMSOWS1, SEQ ID NO:85.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnery;
KW cell culture; chemotaxis; food additive; chromosome X;
KW binding partner identification.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX Peptide 1..19
XX /label= signal_peptide
XX Protein 20..399
XX /note= "Mature secreted protein"
XX
XX WO200136432-A2.
XX
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-US31162.
XX
XX 19-NOV-1999; 99US-0166415.
XX 30-JUN-2000; 2000US-0215136.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;
XX
XX WPI; 2001-343793/36.
XX DR N-PSDB; AAD08520.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition
PT
XX
XX Claim 11; Page 460-461; 509pp; English.
XX
XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.
CC AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 18 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.
XX
XX

```

CC Antibodies specific for a protein of the invention can be used in  
 CC alleviating symptoms associated with the disorders mentioned above, and  
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
 CC immunosorbent assay (ELISA). The present sequence represents a human  
 CC secreted protein of the invention.

XX  
 SQ Sequence 399 AA;

Query Match 92.9%; Score 870; DB 22; Length 399;  
 Best Local Similarity 64.7%; Pred. No. 5e-76;  
 Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

Qy 1 MGILLGILLGLHGLTVDYGRPILEVPESTGPKGDVNLPCYDPLQGYTQVLVKNLQYR 60  
 Db 1 MGILLGILLGLHGLTVDYGRPILEVPESTGPKGDVNLPCYDPLQGYTQVLVKNLQYR 60

Qy 61 GSDPVTIFLRDSSGDHIOQAKYQGRHLVSHKVPDVSLSLSTLEMDRSHYTCVETWTP 120  
 Db 61 GSDPVTIFLRDSSGDHIOQAKYQGRHLVSHKVPDVSLSLSTLEMDRSHYTCVETWTP 120

Qy 121 DGNQVVRDKITELRVQ----- 136  
 Db 121 DGNQVVRDKITELRVQ----- 136

Qy 137 -----KHSSKLLKTK 146  
 Db 137 -----KHSSKLLKTK 146

Qy 147 TEAPTTMTYPLKATSTVKQSDWNTDMDG 175  
 Db 147 TEAPTTMTYPLKATSTVKQSDWNTDMDG 175

Qy 241 TEAPTTMTYPLKATSTVKQSDWNTDMDG 269  
 Db 241 TEAPTTMTYPLKATSTVKQSDWNTDMDG 269

RESULT 13  
 ID ABG64487  
 XX ABG64487 standard; Protein; 399 AA.  
 AC ABG64487;  
 DT 27-AUG-2002 (first entry)  
 XX Human albumin fusion protein #1163.  
 DE Human albumin fusion protein #1163.  
 KW Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KW human serum albumin; HSA; cancer; reproductive disorder;  
 KW digestive disorder; immune disorder; endocrine disorder;  
 KW haematopoietic disorder; neural disorder; connective disorder;  
 KW cytostatic; antiinfertility; antiinflammatory; antiulcer;  
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;  
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritic.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX WO200177137-A1.  
 XX 18-OCT-2001.  
 XX 12-APR-2001; 2001WO-US11988.  
 XX 12-APR-2000; 2000US-229358P.  
 XX 25-APR-2000; 2000US-199384P.  
 XX 21-DEC-2000; 2000US-256931P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Haseltine WA;  
 XX WPI; 2002-010886/01.  
 XX New fusion protein for treating disease e.g. diabetes comprises an  
 PT albumin fused to a therapeutic protein -

XX  
 PS Claim 1; Page 1257-1258; 2102pp; English.  
 XX

CC The present invention relates to albumin fusion proteins comprising a  
 CC therapeutic protein X and human albumin (HA, also known as human serum  
 CC albumin, HSA). The proteins are useful for treating a disease or  
 CC disorder that may be modulated by therapeutic protein X. The albumin  
 CC extends the shelf-life of protein X, and may increase its biological  
 CC in vitro/in vivo activity. The protein is useful for treating and  
 CC diagnosing disorders such as cancer, reproductive disorders, digestive  
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders  
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders  
 CC (e.g. diabetes), haematopoietic disorders, neural disorders  
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,  
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders  
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin  
 CC fusion proteins of the invention.

XX  
 SQ Sequence 399 AA;

Query Match 92.9%; Score 870; DB 23; Length 399;  
 Best Local Similarity 64.7%; Pred. No. 5e-76;  
 Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

Qy 1 MGILLGILLGLHGLTVDYGRPILEVPESTGPKGDVNLPCYDPLQGYTQVLVKNLQYR 60  
 Db 1 MGILLGILLGLHGLTVDYGRPILEVPESTGPKGDVNLPCYDPLQGYTQVLVKNLQYR 60

Qy 61 GSDPVTIFLRDSSGDHIOQAKYQGRHLVSHKVPDVSLSLSTLEMDRSHYTCVETWTP 120  
 Db 61 GSDPVTIFLRDSSGDHIOQAKYQGRHLVSHKVPDVSLSLSTLEMDRSHYTCVETWTP 120

Qy 121 DGNQVVRDKITELRVQ----- 136  
 Db 121 DGNQVVRDKITELRVQ----- 136

Qy 137 -----KHSSKLLKTK 146  
 Db 137 -----KHSSKLLKTK 146

Qy 147 TEAPTTMTYPLKATSTVKQSDWNTDMDG 175  
 Db 147 TEAPTTMTYPLKATSTVKQSDWNTDMDG 175

Qy 241 TEAPTTMTYPLKATSTVKQSDWNTDMDG 269  
 Db 241 TEAPTTMTYPLKATSTVKQSDWNTDMDG 269

RESULT 14  
 ID ABG64488  
 XX ABG64488 standard; Protein; 399 AA.  
 AC ABG64488;  
 DT 27-AUG-2002 (first entry)  
 XX Human albumin fusion protein #1163.

XX Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KW human serum albumin; HSA; cancer; reproductive disorder;  
 KW digestive disorder; immune disorder; endocrine disorder;  
 KW haematopoietic disorder; neural disorder; connective disorder;  
 KW cytostatic; antiinfertility; antiinflammatory; antiulcer;  
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;  
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritic.

XX Homo sapiens.  
 OS Synthetic.  
 XX WO200177137-A1.  
 XX 18-OCT-2001.  
 XX 12-APR-2001; 2001WO-US11988.



```

PR 12-APR-2000; 2000US-229358P.
PR 25-APR-2000; 2000US-199384P.
PR 21-DEC-2000; 2000US-256931P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Haseltine WA;
XX WPI; 2002-010886/01.
XX
XX New fusion protein for treating disease e.g. diabetes comprises an
XX albumin fused to a therapeutic protein -
XX
XX Claim 1; Page 1258-1259; 2102pp; English.
XX
XX The present invention relates to albumin fusion proteins comprising a
XX therapeutic protein X and human albumin (HA, also known as human serum
XX albumin, HSA). The proteins are useful for treating a disease or
XX disorder that may be modulated by therapeutic protein X. The albumin
XX extends the shelf-life of protein X, and may increase its biological
XX in vitro/in vivo activity. The protein is useful for treating and
XX diagnosing disorders such as cancer, reproductive disorders, digestive
XX disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
XX (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
XX (e.g. diabetes), haematopoietic disorders, neural disorders
XX (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
XX encephalomyelitis, meningitis, schizophrenia), and connective disorders
XX (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
XX fusion proteins of the invention.
XX
XX Sequence 399 AA;
XX
XX Query Match 92.9%; Score 870; DB 23; Length 399;
XX Best Local Similarity 64.7%; Pred. No. 5e-76;
XX Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;
XX
XX QY 1 MGILLGLLLGLHLTVDTYGRPILEVPESTVGPKGVDVNLPCYDPLQGYTVLVKWLVR 60
XX DB 1 MGILLGLLLGLHLTVDTYGRPILEVPESTVGPKGVDVNLPCYDPLQGYTVLVKWLVR 60
XX
XX QY 61 GSDPTVIFLRDSSGHHIOAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVETWQFP 120
XX DB 61 GSDPTVIFLRDSSGHHIOAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVETWQFP 120
XX
XX QY 121 DGNVVRDKITELRVQ----- 136
XX DB 121 DGNVVRDKITELRVQ----- 136
XX
XX QY 137 -----KHSSKLLTK 146
XX DB 137 -----KHSSKLLTK 146
XX
XX QY 147 TEAPTMTYPLKATSTVKQSDWMTTMDG 175
XX DB 241 TEAPTMTYPLKATSTVKQSDWMTTMDG 269
XX
XX RESULT 15
XX AAE04290
XX ID AAE04290 standard; Protein; 386 AA.
XX
XX AC AAE04290;
XX
XX DT 09-AUG-2001 (first entry)
XX
XX DE Human gene 11 encoded secreted protein fragment, SEQ ID NO:154.
XX
XX KW Human; secreted protein; proliferative disorder; cancer; tumour;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX inflammation; allergy; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;
XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;

```

```

KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnery;
KW cell culture; chemotaxis; food additive;
KW binding partner identification; chromosome X.
XX
XX Homo sapiens.
XX
XX WO200136432-A2.
XX
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-US31162.
XX
XX 19-NOV-1999; 99US-0166415.
XX 30-JUN-2000; 2000US-0215136.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;
XX WPI; 2001-343793/36.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; Page 504-505; 509pp; English.
XX
XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
XX protein genes, and AAE04199-AAE04239 represent the proteins they encode.
XX AAE04240-AAE04297 represent human secreted protein fragments or variants.
XX
XX The secreted proteins and their genes are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 18 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of
XX proliferative disorders, cancer, tumours, foetal and developmental
XX abnormalities, haematopoietic disorders, diseases of the immune system,
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX allergies, neurological disorders (e.g., Alzheimer's disease,
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX cardiovascular disorders, angiogenic disorders, kidney disorders,
XX gastrointestinal disorders, pregnancy-related disorders, endocrine
XX disorders, and infections. The proteins can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin ageing due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues, to identify their
XX cognate ligands or binding partners, and in chemotaxis, and can be used
XX as a food additive or preservative to modify storage properties.
XX Antibodies specific for a protein of the invention can be used in
XX alleviating symptoms associated with the disorders mentioned above, and
XX in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX immunosorbent assay (ELISA). The present sequence represents a human
XX secreted protein fragment referred to in the disclosure of the invention.
XX
XX Sequence 386 AA;
XX
XX Query Match 86.2%; Score 807; DB 22; Length 386;
XX Best Local Similarity 62.9%; Pred. No. 6.3e-70;
XX Matches 161; Conservative 0; Mismatches 1; Indels 94; Gaps 1;
XX
XX QY 14 TVDTYGRPILEVPESTVGPKGVDVNLPCYDPLQGYTVLVKWLVRGSDPTVIFLRDSS 73
XX DB 1 TVDTYGRPILEVPESTVGPKGVDVNLPCYDPLQGYTVLVKWLVRGSDPTVIFLRDSS 60
XX
XX QY 74 GDHIOAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVETWQFPDGNVVRDKITEL 133
XX DB 61 GDHIOAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVETWQFPDGNVVRDKITEL 120
XX
XX QY 134 RVQ----- 136

```

Db 121 RVOKLSVSKPTVTGSGYGTVPQGMRIISLQOARGSPPISYIWKOOTNNQEPKIVATL 180  
Qy 137 -----KHSSKLLKTKTEAPTMTYPLKA 159  
Db 181 STLLFKPAVIADSGSYFCTAKGVGSEQHSQDIVKFVVKDSKLLKTKTEAPTMTYPLKA 240  
Qy 160 TSTVKQSWDWTMDG 175  
Db 241 TSTVKQSWDWTMDG 256

Search completed: March 4, 2003, 14:58:47  
Job time : 17.0933 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: March 4, 2003, 14:53:52 ; Search time 7.87247 seconds  
(without alignments)  
2137.006 Million cell updates/sec

Title: US-09-763-902B-6  
Perfect score: 936  
Sequence: 1 MGILLGULLLGHLLTVDTYGR.....PLKATSTVRKQSDWTTMDMG 175

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 106.5 | 11.4        | 354    | 1 S42938 | proteoglycan link  |
| 2          | 104.5 | 11.2        | 365    | 2 JC7780 | coxsackie- and ade |
| 3          | 100.5 | 10.7        | 355    | 1 LKCH   | proteoglycan link  |
| 4          | 100.5 | 10.7        | 408    | 1 LKRT2  | proteoglycan link  |
| 5          | 100   | 10.7        | 354    | 1 LKHU   | proteoglycan link  |
| 6          | 97.5  | 10.4        | 144    | 2 S36308 | T-cell receptor de |
| 7          | 97.5  | 10.4        | 354    | 1 S04243 | proteoglycan link  |
| 8          | 97.5  | 10.4        | 453    | 2 B83380 | hypothetical prote |
| 9          | 97.5  | 10.4        | 478    | 2 I53960 | PRR2 alpha - human |
| 10         | 97.5  | 10.4        | 526    | 2 A37821 | butyrophilin - bov |
| 11         | 97.5  | 10.4        | 538    | 2 I68093 | PRR2 delta - human |
| 12         | 97    | 10.4        | 116    | 2 D24891 | T-cell receptor al |
| 13         | 96    | 10.3        | 526    | 2 S70587 | butyrophilin precu |
| 14         | 95    | 10.1        | 137    | 2 I46628 | rearranged T-cell  |
| 15         | 94.5  | 10.1        | 135    | 2 S36314 | T-cell receptor de |
| 16         | 94    | 10.0        | 118    | 2 I71934 | MHC class II I-A-a |
| 17         | 93.5  | 10.0        | 142    | 2 S36307 | T-cell receptor de |
| 18         | 92.5  | 9.9         | 144    | 2 S36322 | T-cell receptor de |
| 19         | 92.5  | 9.9         | 1643   | 2 T14274 | versican precursor |
| 20         | 92.5  | 9.9         | 3381   | 2 T42389 | versican precursor |
| 21         | 92    | 9.8         | 116    | 2 E24891 | T-cell receptor al |
| 22         | 92    | 9.8         | 132    | 2 A53410 | Ig light chain v r |
| 23         | 92    | 9.8         | 309    | 2 T49522 | gene B7-2 protein  |
| 24         | 91.5  | 9.7         | 142    | 2 S04664 | T-cell receptor de |
| 25         | 91    | 9.7         | 531    | 2 S20900 | titin - mouse (fra |
| 26         | 89.5  | 9.6         | 323    | 2 A48997 | tumor surface anti |
| 27         | 89    | 9.5         | 101    | 2 A28165 | proteoglycan link  |
| 28         | 89    | 9.5         | 145    | 2 I46629 | rearranged T-cell  |
| 29         | 89    | 9.5         | 299    | 2 S56749 | junctional adhesio |

30 89 9.5 340 2 T28137 Ig V-region-like B  
31 89 9.5 984 2 A55137 hyaluronate lyase  
32 88.5 9.5 7962 2 I38346 elastic titin - hu  
33 88 9.4 205 2 A48929 activated B-cell p  
34 88 9.4 247 2 A55717 myelin/oligodendro  
35 87.5 9.3 246 2 A47712 T-cell receptor de  
36 87.5 9.3 293 2 A40131 probable purine nu  
37 87.5 9.3 1518 2 S37928 T-cell receptor de  
38 87 9.3 139 2 S36302 secretory componen  
39 87 9.3 773 1 ORRBG titin, cardiac mus  
40 87 9.3 26926 1 I38344 T-cell receptor al  
41 86.5 9.2 115 2 C24891 Ig V-region-like B  
42 86.5 9.2 372 2 C39371 T-cell receptor al  
43 86 9.2 116 2 B24891 Ig lambda chain -  
44 85.5 9.1 243 2 S25755 T-cell receptor de  
45 85.5 9.1 292 2 S03421 T-cell receptor de

ALIGNMENTS

RESULT 1

S42938  
proteoglycan link protein precursor - horse  
N: Alternate names: cartilage link protein  
C: Species: Equus caballus (domestic horse)  
C: Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 22-Jun-1999  
C: Accession: S42938  
R: Dudhia, J.; Platt, D.  
submitted to the EMBL Data Library, March 1994  
A: Reference number: S42938  
A: Accession: S42938  
A: Molecule type: mRNA  
A: Residues: 1-354 <DUD>  
A: Cross-references: EMBL: X78077; NID: g459438; PIDN: CAA54987.1; PID: g459439  
C: Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repea  
C: Keywords: cartilage; duplication; extracellular matrix; glycoprotein  
F: 1-15/Domain: signal sequence #status predicted <SIG>  
F: 16-354/Product: proteoglycan link protein #status predicted <MAT>  
F: 54-141/Domain: immunoglobulin homology <IMM>  
F: 176-253/Domain: link protein repeat homology <LINK1>  
F: 274-350/Domain: link protein repeat homology <LINK2>  
F: 21-56/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F: 61-139/Disulfide bonds: #status predicted

Query Match 11.4%; Score 106.5; DB 1; Length 354;  
Best Local Similarity 29.8%; Pred. No. 0.089;  
Matches 37; Conservative 19; Mismatches 57; Indels 11; Gaps 6;

QY 12 HLTVDYGPPILEVPES-VTGPNGVDNLPCTY--DPL--QGYTVLVKW--LVQRGSD '63  
DB 31 HIQAENGPRLLVEAEQKVFSHRGGNVTLPCKFLRDPFAFGSGTHKIRIKTKLSDYLK '90  
QY 64 PVTIFLRDSSGGHIO-QAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVETWQTDPG '122  
DB 91 EYDVFV--SMGYHKYTYGQGRVFLKGGSDNDASLVITDLDLDDYGRYKCEVIGLEDD '148  
QY 123 NQVW 126  
DB 149 TAVV 152

RESULT 2

JC7780  
coxsackie- and adenovirus receptor - bovine  
C: Species: Bos primigenius taurus (cattle)  
C: Date: 02-Apr-2002 #sequence\_revision 02-Apr-2002 #text\_change 02-Apr-2002  
C: Accession: JC7780  
R: Thoenen, I.; Keyaerts, E.; Lindberg, M.; Van Ranst, M.  
Biochem. Biophys. Res. Commun. 288, 805-808, 2001  
A: Title: Characterization of a cDNA encoding the bovine coxsackie and adenovirus rece  
A: Reference number: JC7780  
A: Contents: Liver

A:Accession: JCT780

A:Molecule type: mRNA

A:Residues: 1-365 <THO>

A:Cross-references: GB:AY033651

C:Comment: This protein serves as the primary adenoviral attachment site on bovine cells

Query Match 11.2%; Score 104.5; DB 2: Length 365;

Best Local Similarity 27.6%; Pred. No. 0.14;

Matches 56; Conservative 25; Mismatches 81; Indels 41; Gaps 12;

QY 1 MGLLGLLLGLLHGLTVDYGRPILVPEVSGPWKGD-VNLPCY-----DPLQGYTQV 52

Db 1 MELLRLFLLCGVADFTGLSI-TTPEQMIKAGETAYLPCKFTGLPQDQPLD----- 54

QY 53 LVKWLIVQSGS---DPVTFELRSSGDHIQAKYQ---CRLHVSFK--VPGDVSILQLSTL 103

Db 55 -LEWLLSPADNKVKVQVILY---SGDKLYDYQDLKGRVHFTSNDLKSGDASINVTNL 110

QY 104 EMDRSHYTCVYVTPDGNQ---VVRDKITELRVQRHSSKLL---KTKTEAPTWTY 155

Db 111 QLSIDIGTYOCKVKKAPGVGNKKIQLTVLVKPSGIRCYVDGSEIGNDFLKCE-PKESGL 169

QY 156 PL-----KATSVKQSDWTDM 173

Db 170 PURYEQWQLSDSQKLPTSNLPEM 192

RESULT 3

LACH

Proteoglycan link protein precursor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 22-Jun-1999

C:Accession: A28305; A24881; A39097; B39097

R:Kiss, I.; Deak, F.; Mestric, S.; Delius, H.; Soos, J.; Dekany, K.; Argyraves, W.S.; Spa

Proc. Natl. Acad. Sci. U.S.A. 84, 6399-6403, 1987

A:Title: Structure of the chicken link protein gene: exons correlate with the protein do

A:Reference number: A28305; MUID:87317659; PMID:3476955

A:Accession: A28305

A:Molecule type: DNA

A:Residues: 1-355 <KIS>

A:Cross-references: GB:M35038; NID:g212264; PIDN:AAA48941.1; PID:g212267

R:Deak, F.; Kiss, I.; Sparks, K.J.; Argyraves, W.S.; Hamplkian, G.; Goetinck, P.F.

Proc. Natl. Acad. Sci. U.S.A. 83, 3766-3770, 1986

A:Title: Complete amino acid sequence of chicken cartilage link protein deduced from cDN

A:Reference number: A24881; MUID:86233315; PMID:3459154

A:Accession: A24881

A:Molecule type: protein

A:Residues: 1-355 <DEA>

A:Cross-references: GB:M13212; NID:g212259; PIDN:AAA48940.1; PID:g212260

A:Experimental source: embryonic sternal cartilage

R:Wu, L.N.Y.; Genge, B.R.; Wuthier, R.E.

J. Biol. Chem. 266, 1187-1194, 1991

A:Title: Association between proteoglycans and matrix vesicles in the extracellular matr

A:Reference number: A39097; MUID:9109230; PMID:1985942

A:Accession: A39097

A:Molecule type: protein

A:Residues: 40-55; 'X', 57-60; 'X', 62-75; 'X', 77-78; 'X' <WUA>

A:Note: 38K protein, a major component of matrix vesicles; Asn-56 appears to be glycosyl

A:Accession: B39097

A:Molecule type: protein

A:Residues: 40-55; 'X', 57-59 <WU2>

A:Note: 39K protein, a minor component of matrix vesicles

C:Comment: Link proteins stabilize the aggregates of proteoglycan monomers with hyaluron

structures that induce mineralization).

C:Genetics:

A:Introns: 34/1; 159/1; 260/1

A:Note: single copy gene

C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat h

C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-355/Product: link protein #status predicted <MAT>

F:54-142/Domain: immunoglobulin homology <IMM>

F:177-254/Domain: link protein repeat homology <LNK1>

F:275-351/Domain: link protein repeat homology <LNK2>

F:21/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:56/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:161-140,182-253,206-227,280-350,305-326/Disulfide bonds: #status predicted

Query Match

Best Local Similarity 10.7%; Score 100.5; DB 1: Length 355;

Matches 31; Conservative 14; Mismatches 45; Indels 11; Gaps 4;

QY 35 GDVNLPC-----TYDPLQGYTQVLVKW--LVQRGSDPVTFELRSSGDHIQO-AKYQGR 85

Db 55 GNVTLPCRFYHEHTSTAGSGTHKIRKTKLSTDLKEDVDFV--AMGHRKSKYKYQGR 112

QY 86 LHVSHKVPEDVSLSQSTLEMDRSHYTCVYVTPDGNQV 126

Db 113 VFLRESSENDASLIITNIMLEDYGRYKVEIEGLEDDTAVV 153

RESULT 4

LKRT2

Proteoglycan link protein 2 precursor - rat

N:Alternate names: cartilage link protein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 13-Aug-1986 #sequence\_revision 24-Oct-1997 #text\_change 13-Nov-1998

C:Accession: A28654; A24880; A02869

R:Rhodes, C.; Doerge, K.; Sasaki, M.; Yamada, Y.

J. Biol. Chem. 263, 6063-6067, 1988

A:Title: Alternative splicing generates two different mRNA species for rat link prote

A:Reference number: A28654; MUID:88198139; PMID:2452158

A:Accession: A28654

A:Molecule type: mRNA

A:Residues: 1-408 <RHO>

R:Doerge, K.; Hassell, J.R.; Catterson, B.; Yamada, Y.

Proc. Natl. Acad. Sci. U.S.A. 83, 3761-3765, 1986

A:Title: Link protein cDNA sequence reveals a tandemly repeated protein structure.

A:Reference number: A24880; MUID:86233314; PMID:3459153

A:Accession: A24880

A:Molecule type: protein

A:Residues: 16-33; 'A', 89-375; 'W', 377-408 <NEA>

C:Comment: This protein was extracted from rat chondrosarcoma.

C:Comment: Residues 259-277 and 358-376 (approximately) form epitopes for the species

C:Superfamily: Link proteins interact with and stabilize aggregates of hyaluronic acid an

C:Keywords: proteoglycan link protein; immunoglobulin homology; link protein repea

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-408/Product: proteoglycan link protein 2 #status experimental <MAT>

F:108-195/Domain: immunoglobulin homology <IMM>

F:230-307/Domain: link protein repeat homology <LNK1>

F:328-404/Domain: link protein repeat homology <LNK2>

F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:115-193,235-306,259-280,333-403,358-379/Disulfide bonds: #status experimental

Query Match

Best Local Similarity 10.7%; Score 100.5; DB 1: Length 408;

Matches 38; Conservative 21; Mismatches 55; Indels 17; Gaps 7;

QY 8 LLLGLHGLTVDYGRPILVPEVSGPWKGDVNLPCY--DPL---QGYTQVLVKWLVRG 61

Db 81 LATSDRTTEGPRLLVVEAQKVFSGHGGNVTLPCKFYRDPFTAFSGGIHKIRIKW-TKLT 139

QY 62 SDPVTFILRD-----SSGDHIQ-QAKYQGRHSHKVPDVSLSQSTLEMDRSHYTCV 115

Db 140 SD----YLREVDVFSVNGYHKYGGYQGRVFLKSGSDNDASLIITDLTLEDYGRYKVE 195

QY 116 TWOTPDGNQV 126

Db 196 IEGLDPTNAV 206

## RESULT 5

LKRU

proteoglycan link protein precursor [validated] - human  
N:Alternate names: cartilage link protein

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1991 #sequence, revision 31-Dec-1991 #text, change 08-Dec-2000  
C:Accession: S14914; S04244; S03868; A36308; S14926; S09309; S08041

R:Dudhia, J.; Hardingham, T.E.

Nucleic Acids Res. 18, 1292, 1990

A:Title: The primary structure of human cartilage link protein.

A:Reference number: S14914; MUID:90206798; PMID:2320422

A:Accession: S14914

A:Molecule type: mRNA

A:Residues: 1-354 <DUD>

A:Cross-references: EMBL:X17405; NID:q463246; PIDN:CA35462.1; PID:g34378

A:Note: the authors translated the codon GAT for residue 93 as Ala

R:Dudhia, J.; Hardingham, T.E.

J. Mol. Biol. 206, 749-753, 1989

A:Title: Appendix. Isolation and sequence of cDNA clones for pig and human cartilage link

A:Reference number: S04243

A:Accession: S04244

A:Molecule type: mRNA

A:Residues: 223-354 <DUD>

A:Cross-references: EMBL:Y00166

A:Note: the authors translated the codon CTG for residue 264 as Arg and TNG for residue

Blochem. J. 259, 61-67, 1989

A:Title: Degradation of proteoglycan aggregate by a cartilage metalloproteinase. Evidence

A:Reference number: S03868; MUID:89246328; PMID:2719651

A:Accession: S03868

A:Molecule type: protein

A:Residues: 16-35 <NGU>

R:Osborne-Lawrence, S.L.; Sinclair, A.K.; Hicks, R.C.; Lacey, S.W.; Eddy Jr., R.L.; Byer

Genomics 8, 562-567, 1990

A:Title: Complete amino acid sequence of human cartilage link protein (CRTLI) deduced fr

A:Reference number: A36308; MUID:91139126; PMID:2286376

A:Accession: A36308

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra

A:Molecule type: mRNA

A:Residues: 1-354 <OSB>

R:Dudhia, J.; Hardingham, T.E.

Nucleic Acids Res. 18, 2214, 1990

A:Title: The primary structure of human cartilage link protein.

A:Reference number: S14926; MUID:90245703; PMID:2336413

A:Accession: S14926

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-92, 'A', '94-354 <NUC>

R:Nguyen, O.; Mort, J.S.; Roughley, P.J.

Blochem. J. 266, 569-573, 1990

A:Title: Cartilage proteoglycan aggregate is degraded more extensively by cathepsin L th

A:Reference number: S09309; MUID:90197639; PMID:2317204

A:Accession: S09309

A:Molecule type: protein

A:Residues: 16-38, 'X', '40-55, 57-60, 62-65, 'X', '67, 'X', '69-80 <NG2>

C:Genetics:

A:Gene: GDB:CR11

A:Cross-references: GDB:125232; OMIM:115435

A:Map position: 5q13-5q14

C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat

C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-35/Product: proteoglycan link protein #status experimental <MAT>

F:34-141/Domain: immunoglobulin homology <IMM>

F:16-253/Domain: link protein repeat homology <LNK1>

F:274-350/Domain: link protein repeat homology <LNK2>

F:21-56/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:61-139/Disulfide bonds: #status predicted

Query Match 10.7%; Score 100; DB 1; Length 354;  
Best Local Similarity 32.0%; Pred. No. 0.33; Mismatches 44; Indels 10; Gaps 5;  
Matches 32; Conservative 14; Mismatches 44; Indels 10; Gaps 5;

## Query

35 GDVNLPCY--DPL---QGYQVAVK--LVQRGSDPYTIFLRDSDGHIO-QAYQGR 86

Db 55 GNVLPCKFYRDPNPAFGGIRKIKTKTKLSDLKLEVDV--SMGYHKTKYGGYGRV 112

## Query

87 HVSHKVPDVSQSLSTLEMDRSHYTCVWTQTPDGNQV 126

Db 113 FLKGSDDASLAVITDILLEDYGRKCEVIEGLBDPVMV 152

## RESULT 6

S36308

T-cell receptor delta chain precursor - sheep (fragment)

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 03-Feb-1994 #sequence, revision 03-Feb-1994 #text, change 23-Jul-1999

C:Accession: S36308; S23024

R:Hein, W.R.; Dudler, L.

EMBO J. 12, 715-724, 1993

A:Title: Divergent evolution of T cell repertoires: extensive diversity and developme

A:Reference number: S36287; MUID:93178447; PMID:8440261

A:Accession: S36308

A>Status: preliminary; translation not shown

A:Molecule type: mRNA

A:Residues: 1-144 <HEI>

A:Cross-references: EMBL:Z12980; NID:92225; PIDN:CA47324.1; PID:g2226

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: T-cell receptor

Query Match 10.4%; Score 97.5; DB 2; Length 144;  
Best Local Similarity 26.7%; Pred. No. 0.18;

Matches 27; Conservative 16; Mismatches 45; Indels 13; Gaps 4;

Query 37 VNLCTFDPLGCTQVAVKMLVQRGSDPYTIFLRDSDGHIO-QAYQGR 95

Db 39 ITLWCRFEASDMSTYIFWYKQLPSGKMTLLHQHSDGNARGRSVNFQAKH---S 95

Query 96 VSLQSLSTLEMDRSHYTCVWT---NOTPDGNQVVDKIN 131

Db 96 ISLTISALIELDQSAKRYCVLSYGIW---GTLTSDKLT 132

## RESULT 7

S04243

proteoglycan link protein precursor - pig

N:Alternate names: cartilage link protein

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 28-Feb-1990 #sequence, revision 28-Feb-1990 #text, change 22-Jun-1999

C:Accession: S04243; I47145

R:Dudhia, J.; Hardingham, T.E.

J. Mol. Biol. 206, 749-753, 1989

A:Title: Appendix. Isolation and sequence of cDNA clones for pig and human cartilage

A:Reference number: S04243

A:Accession: S04243

A:Molecule type: mRNA

A:Residues: 1-354 <DUD>

A:Cross-references: EMBL:Y00165; NID:92009; PIDN:CA468358.1; PID:g2010

R:Perkins, S.J.; Nealis, A.S.; Dudhia, J.; Hardingham, T.E.

J. Mol. Biol. 206, 737-753, 1989

A:Title: Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal

A:Reference number: I47145; MUID:89293837; PMID:2738916

A:Accession: I47145

A>Status: preliminary; translated from GB/EMBL/DBI

A:Molecule type: mRNA

A:Residues: 1-354 <PER>

A:Cross-references: EMBL:Y00165; NID:92009; PIDN:CA468358.1; PID:g2010

C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repea

C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-35/Product: proteoglycan link protein #status predicted <MAT>

F:34-141/Domain: immunoglobulin homology <IMM>

F:176-253/Domain: link protein repeat homology <LNK1>  
F:274-350/Domain: link protein repeat homology <LNK2>  
F:21.56/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:61-139/Disulfide bonds: #status predicted

Query Match 10.4%; Score 97.5; DB 1; Length 354;  
Best Local Similarity 28.2%; Pred. No. 0.55;  
Matches 35; Conservative 20; Mismatches 58; Indels 11; Gaps 6;

QY 12 HLTVDYGRPILEVPEV-VTGPWKGDVNLPCY--DPL---QGYQVLVKV--LVQRGSD 63  
DB 31 HQAENGPRLLVEAQKAFVSRGGNTVLPCKFFRDPTAFSGTHKIRIKTKLSDYLK 90  
QY 64 PVTIFLRDSSGHQIQ-QAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVETWTQTPDG 122  
DB 91 EYDVFV--SMGYHKYGYGHRVFLKGGSDNDASLIVTDLTDYGRYKCEVIEGLEDD 148  
QY 123 NOVW 126  
DB 149 TAVV 152

RESULT 8  
B83380  
Hypothetical protein PA2131 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83380  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lozy, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: B83380  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-453 <STO>  
A:Cross-references: GB:AE004640; GB:AE004091; NID:9948139; PIDN:AG05519.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2131

Query Match 10.4%; Score 97.5; DB 2; Length 453;  
Best Local Similarity 21.7%; Pred. No. 0.75;  
Matches 50; Conservative 26; Mismatches 57; Indels 87; Gaps 9;

QY 1 MGILLGLLLGLHLTV-----DTYGRPILEVPEVSTGPW-----KGDVNL 40  
DB 13 LGVLLGLPASQATCYRVTSVSGSETTSNTQIRP-----GEGTAGSWTGACDTCNGSLGLP 68  
QY 41 C-----TYDPLOGYQVLVKVLVQRGS-----DPVTIFLRDSSGDHIQQA-----KY 82  
DB 69 SVNVNDSFQPYPTLIASAVAPLTQYGATGGYDPERVFCRSAGDAVYEMFTNGDDLY 128  
QY 83 QGRHLVSHKVPDGVSL-----QLSTLEMDRSHY 111  
DB 129 SGYTGADTVGNDIGLDQAVATAMPVLLRLTNVETGTTIPNKRQLSGLDISR--- 185

QY 112 TCEVTWTPDGNQVVRDK-ITELVRQKSHSKLLTKTKTEAPTMTYPLKAT 160  
DB 186 -----F--GFQLVAKNLAARVLYSAPLEAIRVYSPTILSQPSY 225

RESULT 9  
153960  
PRR2 alpha - human  
C:Species: Homo sapiens (man)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 21-Jan-2000  
C:Accession: 153960  
R:Eberle, F.; Dubreuil, P.; Mattei, M.G.; Devillard, E.; Lopez, M.  
Gene 159, 267-272, 1995  
A:Title: The human PRR2 gene, related to the human poliovirus receptor gene (PVR), is th

A:Reference number: 153960; MUID:95347610; PMID:7622062  
A:Accession: 153960  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-478 <RES>  
A:Cross-references: GB:S79171; NID:g1042202; PID:g1042203  
C:Genetics:  
A:Gene: PRR2alpha  
C:Superfamily: poliovirus receptor; immunoglobulin homology  
F:276-331/Domain: immunoglobulin homology <IMM>

Query Match 10.4%; Score 97.5; DB 2; Length 478;  
Best Local Similarity 25.7%; Pred. No. 0.8;  
Matches 52; Conservative 28; Mismatches 75; Indels 47; Gaps 11;

QY 3 ILGLLLGLHLTVDTYGRPI-LEVPESVTGPMKGDVNLPC-TYDPLOGYQVLVKVLVQR 60  
DB 17 LLWPLLLL--LLETGAQDVRVOVLPEVRGQLGGTVLPCLLPPVGLYISLVTV--QR 72  
QY 61 GSDPVTIFLRDSSGDHIQAKYQGRHLVSHKVP-----GDV 96  
DB 73 -----PDAPANHQNVAAPFKMGPSFPSPKPGSERLSVSAKQSGQDTEALQDA 123

QY 97 SIQSLTLEMDRSHYTCVETWTPDGNQVVRDKITELRV---QKHSSKLLK-TKTEAPT 152  
DB 124 TLALHGLTVDEGNTCEFA-TFPKGS--VRG-MWLVRIAKPKNAQAEAKVTFSDPTT 179

QY 153 MTPYPLKATSVKQSWDWTMD 174  
DB 180 VALCISKEGPPARISWLSLD 201

RESULT 10  
A37821  
butyrophilin - bovine  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 30-Apr-1991 #sequence\_revision 30-Apr-1991 #text\_change 05-Nov-1999  
C:Accession: A37821  
R:Jack, L.J.W.; Mather, I.H.  
J. Biol. Chem. 265, 14481-14486, 1990  
A:Title: Cloning and analysis of cDNA encoding bovine butyrophilin, an apical glycopr

A:Reference number: A37821; MUID:90354441; PMID:2387867  
A:Accession: A37821  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-526 <JAC>  
A:Cross-references: GB:M35551; NID:g1763685; PIDN:AAB39766.1; PID:g162773  
C:Keywords: transmembrane protein

Query Match 10.4%; Score 97.5; DB 2; Length 526;  
Best Local Similarity 24.1%; Pred. No. 0.9;  
Matches 39; Conservative 23; Mismatches 85; Indels 15; Gaps 5;

QY 3 ILGLLLGLHLTVDTYGRPI-LEVPESVTGPMKGDVNLPC-TYDPLOGYQVLVKVLVQR 60  
DB 17 ILQLPLKDSAPFDVIGPQEPIL-----AVVGE---DAELPCLSPNVSAKGMELRWFEK 69

QY 61 GSDPVTIFLRDSSGDHIQAKYQGRHLV--SHKVPDGVSLQSLTLEMDRSHYTCVETW 118  
DB 70 VSPAVFVSREGQEGEGEMAEYGRVSLVEDHIAEGSVAVRIQEVKASDDGDEYRCFFRQD 129

QY 119 TPDGNQVVRDKITELRVQKHSSKLLTKTKTEAPTMT-----YP 156  
DB 130 ENYEEAIVHLKVAALGSDPHISMVKVQESGEIQLECTSVGWYP 171

RESULT 11  
I68093  
PRR2 delta - human  
C:Species: Homo sapiens (man)  
C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jan-2000  
C:Accession: I68093

R:Eberle, F.; Dubreuil, P.; Mattei, M.G.; Devillard, E.; Lopez, M.  
Gene 159, 267-272, 1995  
A:Title: The human PRX2 gene, related to the human poliovirus receptor gene (PVR), is th  
A:Reference number: I53960; MUID:95347610; PMID:7622062  
A:Accession: I58093  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-538 <RES>  
A:Cross-references: GB:S79172; NID:g1042204; PID:g1042205  
C:Genetics:  
A:Gene: PRR2delta  
C:Superfamily: poliovirus receptor; immunoglobulin homology  
F:276-331/Domain: immunoglobulin homology <IMM>

Query Match 10.4%; Score 97.5; DB 2; Length 538;  
Best Local Similarity 25.7%; Pred. No. 0.92; Mismatches 75; Indels 47; Gaps 11;  
Matches 52; Conservative 28;  
QY 3 ILGLLLGLHGLTVDYGRPI-LEVPESVTGPKGVDNLPCL-TPDPLQGYTVLQVWLVOR 60  
DB 17 LWPPLLL--LLETGAQDVRVQVLPVGRGOLGGTVLPCHLLPPVGLYISLVW--QR 72  
QY 61 GSDPTIFLRDSSGHHIOAKYQGRHLVSHKVP-----GDV 96  
DB 73 -----PDAPANHQNVAAPFKPKGSPFPKPGSERLSEVSASQSTGQDTEAELODA 123  
QY 97 SIQLSTLEMDRSHYTCVETWTPDGNQVVRDKINELRV---QKHSKLLK-TKTEAPT 152  
DB 124 TLALHGLVDEGNTCEFA-TFFKGS--VRG-MTWRVIAKPKNOAEAKQVTFSDPT 179  
QY 153 MTPYPLKATSTVKQSDWTDMD 174  
DB 180 VALCISKGRPPARISWSSLD 201

RESULT 12  
D24891  
T-cell receptor alpha chain precursor V region (F3.5) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 21-Jan-2000  
C:Accession: D24891  
R:Chou, H.S.; Behlke, M.A.; Godambe, S.A.; Russell, J.H.; Brooks, C.G.; Loh, D.Y.  
EMBO J. 5, 2149-2155, 1986  
A:Title: T cell receptor genes in an allelically active CTL clone: implications for rearrangem  
A:Reference number: A31048; MUID:87053852; PMID:3490968  
A:Accession: D24891  
A:Molecule type: DNA  
A:Residues: 1-116 <CHO>  
A:Note: this sequence was determined from the germline gene  
A:Note: the authors translated the codon GGA for residue 18 as Arg  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-116/Product: T-cell receptor alpha chain V region F3.5 #status predicted <VAR>  
F:37-114/Domain: immunoglobulin homology <IMM>

Query Match 10.4%; Score 97; DB 2; Length 116;  
Best Local Similarity 27.6%; Pred. No. 0.15;  
Matches 32; Conservative 23; Mismatches 49; Indels 12; Gaps 4;  
QY 3 ILGLLLGLHGLTVD--TYGRPILEVPESVTGPKGVDNLPCTYDPLQGYTVLQVWLVOR 60  
DB 10 VLVLMLLGRSNGDSVTQTEGLVTTTEGL-----PVLNCTYQ--TAYSDVAFFYVQY 61  
QY 61 GSDPTIFLRDSSGHHIOAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVET 116  
DB 62 LNEAPKLLRSSTDN--KXTEHQGHFATLHKSSSFHLOKSSVQLSDSALYCALIS 115

RESULT 13  
S70587  
butyrophilin precursor - human  
C:Species: Homo sapiens (man)

C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 21-Jul-2000  
C:Accession: S70587  
R:Taylor, M.R.; Peterson, J.A.; Ceriani, R.L.; Couto, J.R.  
Biochim. Biophys. Acta 1306, 1-4, 1996  
A:Title: Cloning and sequence analysis of human butyrophilin reveals a potential rece  
A:Reference number: S70587; MUID:96201696; PMID:8611614  
A:Accession: S70587  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-526 <TAY>  
A:Cross-references: EMBL:U39576; NID:g1326082; PIDN:AAC50489.1; PID:g1326083

Query Match 10.3%; Score 96; DB 2; Length 526;  
Best Local Similarity 24.5%; Pred. No. 1.2; Mismatches 77; Indels 10; Gaps 5;  
Matches 37; Conservative 27;  
QY 4 ILGLLLGLHGLTVDYGRPILEVPESVTGPKGVDNLPCTYDPLQGYTVLQVWLVOR 63  
DB 13 LLTLILLQPLKDSAPFDVIGPEPIAVVGEDAELPCRLSPNASHLELWFRKVS- 71  
QY 64 PVTIFLRDSSGHHIOQ-AKYQGRHLVSHK--VPGDVSLSLSTLEMDRSHYTCVETWQTP 120  
DB 72 PAVLVHRDGRGDEAEQMPYRGRATLVQDGIAGRVLRVIRGVSRVSDGGEYTC---FFRE 128  
QY 121 DGN--QVVRDKITELRVOKHSHKLLKTKE 148  
DB 129 DGSYEALVHLKVAALGSDPHISMVQVENCE 159

RESULT 14  
I46628  
rearranged T-cell receptor delta-chain/Vdelta.6-Ddelta.5-Jdelta.1 - pig (fragment)  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 23-Jul-1999  
C:Accession: I46628  
R:Yang, Y.G.; Ohta, S.; Yamada, S.; Shimizu, M.; Takagaki, Y.  
J. Immunol. 155, 1981-1993, 1995  
A:Title: Diversity of T cell receptor delta-chain cDNA in the thymus of a one-month-o  
A:Reference number: I46623; MUID:95363165; PMID:7636249  
A:Accession: I46628  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-137 <YAN>  
A:Cross-references: GB:D49569; NID:g1041140; PIDN:BAA08513.1; PID:g1041141  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: T-cell receptor

Query Match 10.1%; Score 95; DB 2; Length 137;  
Best Local Similarity 27.0%; Pred. No. 0.28; Mismatches 45; Indels 20; Gaps 7;  
Matches 34; Conservative 27;  
QY 17 TYGRPILE--EVPESTGPKGVDNLPCTYDPLQGYTVLQVWLVORSGSDPVIFLRD-SS 73  
DB 25 TDQPVSRQVEEVT-----LNCRYE--TSNXYILFWYKQPPSGEMFLRQYST 74  
QY 74 GHIOQAKYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVETWQTPDG---NQVVRDKI 130  
DB 75 GWNKEDRYISNFORAKQ---SISLTISALOMKDSATYFC-VLWATWTGRWEDKLIFGKG 130  
QY 131 TELRVQ 136  
DB 131 TOLVVE 136

RESULT 15  
S36314  
T-cell receptor delta chain precursor - sheep (fragment)  
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 23-Jul-1999  
C:Accession: S36314; S23029  
R:Hein, W.R.; Dudler, L.  
EMBO J. 12, 715-724, 1993  
A:Title: Divergent evolution of T cell repertoires: extensive diversity and developme

A:Reference number: S36287; MUID:93178447; PMID:8440261.  
A:Accession: S36314  
A>Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-135 <HEI>  
A:Cross-references: EMBL:Z12986; NID:G22336; PIDN:CAA78330.1; PID:G2237  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: T-cell receptor

Query Match 10.1%; Score 94.5; DB 2; Length 135;  
Best Local Similarity 27.7%; Pred. No. 0.3;  
Matches 28; Conservative 20; Mismatches 46; Indels 7; Gaps 4;

QY 37 VNLPCYDPLQGYTQVLVKKLVORGS-DPTIFLRDSSGDHIOQAKYQGLRVSHKVPGD 95

Db 39 VTLSRYETIOSRYNIF--WKQLPSGEMYLIGCGSSONARYGRYSVNLORSRK---S 93

QY 96 VSLQSLTLEMDDRSHYTCVETWOTPDGNQVVRDKITELRVQ 136

Db 94 ISUTISDELEDSAKYFCAL-WELVGQNPLIFGKGTYLNVE 133

Search completed: March 4, 2003, 15:04:50  
Job time : 9.87247 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:51:02 ; Search time 4.18902 Seconds  
(without alignments)  
1732.709 Million cell updates/sec

Title: US-09-763-902B-6

Perfect score: 936

Sequence: 1 MGILLGILLGLHGLTVDTYGR.....PLKATSVKQSDWTDMDG 175

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Match | Length | DB ID | Description |
|------------|-------|-------|--------|-------|-------------|
| 1          | 106.5 | 11.4  | 354    | 1     | PLK_HORSE   |
| 2          | 103.5 | 11.1  | 356    | 1     | PLK_MOUSE   |
| 3          | 102.5 | 11.0  | 319    | 1     | A33_HUMAN   |
| 4          | 102.5 | 11.0  | 354    | 1     | PLK_RAT     |
| 5          | 100.5 | 10.7  | 355    | 1     | PLK_CHICK   |
| 6          | 100.5 | 10.7  | 354    | 1     | PLK_HUMAN   |
| 7          | 97.5  | 10.4  | 354    | 1     | PLK_BOVIN   |
| 8          | 97.5  | 10.4  | 354    | 1     | PLK_PIG     |
| 9          | 97.5  | 10.4  | 526    | 1     | BUTY_BOVIN  |
| 10         | 97.5  | 10.4  | 538    | 1     | PVR2_HUMAN  |
| 11         | 96    | 10.3  | 526    | 1     | BUTY_HUMAN  |
| 12         | 95.5  | 10.2  | 365    | 1     | CXAR_HUMAN  |
| 13         | 92.5  | 9.9   | 245    | 1     | MOG_RAT     |
| 14         | 92.5  | 9.9   | 3381   | 1     | PGCV_BOVIN  |
| 15         | 92    | 9.8   | 309    | 1     | CD86_MOUSE  |
| 16         | 89.5  | 9.6   | 298    | 1     | JAM1_BOVIN  |
| 17         | 89.5  | 9.6   | 323    | 1     | CD47_HUMAN  |
| 18         | 89    | 9.5   | 299    | 1     | JAM1_HUMAN  |
| 19         | 89    | 9.5   | 984    | 1     | HYSA_STRAG  |
| 20         | 88    | 9.4   | 205    | 1     | CD83_HUMAN  |
| 21         | 87.5  | 9.3   | 246    | 1     | MOG_BOVIN   |
| 22         | 87.5  | 9.3   | 515    | 1     | PVR1_PIG    |
| 23         | 87.5  | 9.3   | 1518   | 1     | KKL1_YEAST  |
| 24         | 87    | 9.3   | 246    | 1     | MOG_MOUSE   |
| 25         | 87    | 9.3   | 773    | 1     | PIGR_RABIT  |
| 26         | 86.5  | 9.2   | 365    | 1     | CXAR_MOUSE  |
| 27         | 86.5  | 9.2   | 1914   | 1     | KMLS_HUMAN  |
| 28         | 85.5  | 9.1   | 771    | 1     | PIGR_MOUSE  |
| 29         | 85    | 9.1   | 247    | 1     | MOG_HUMAN   |
| 30         | 85    | 9.1   | 916    | 1     | CAD4_HUMAN  |
| 31         | 85    | 9.1   | 3396   | 1     | PGCV_HUMAN  |
| 32         | 84.5  | 9.0   | 924    | 1     | ICA5_HUMAN  |
| 33         | 84    | 9.0   | 3358   | 1     | PGCV_MOUSE  |

|    |      |     |      |   |            |
|----|------|-----|------|---|------------|
| 34 | 83.5 | 8.9 | 322  | 1 | ICOL_MOUSE |
| 35 | 83.5 | 8.9 | 417  | 1 | PVR_HUMAN  |
| 36 | 83   | 8.9 | 239  | 1 | CD8A_CANFA |
| 37 | 83   | 8.9 | 240  | 1 | CD7_HUMAN  |
| 38 | 83   | 8.9 | 517  | 1 | PVR1_HUMAN |
| 39 | 82.5 | 8.8 | 300  | 1 | JAM1_MOUSE |
| 40 | 82.5 | 8.8 | 524  | 1 | BUTY_MOUSE |
| 41 | 82.5 | 8.8 | 1906 | 1 | KMLS_CHICK |
| 42 | 81.5 | 8.7 | 215  | 1 | CIB2_RAT   |
| 43 | 81.5 | 8.7 | 828  | 1 | BGAL_BRAOL |
| 44 | 80.5 | 8.6 | 142  | 1 | VPR2_MOUSE |
| 45 | 80.5 | 8.6 | 417  | 1 | PVR_CERAE  |

#### ALIGNMENTS

RESULT 1  
PLK\_HORSE  
ID PLK\_HORSE STANDARD; PRT; 354 AA.  
AC Q28381;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Proteoglycan link protein precursor (Cartilage link protein) (LP).  
GN CTR11.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID-9796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Cartilage;  
RX MEDLINE-96039874; PubMed-7574168;  
RA Dudhia J., Platt D.;  
RT "Complete primary sequence of equine cartilage link protein deduced from complementary DNA".  
RL Am. J. Vet. Res. 56:959-965(1995).  
CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX (BY SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X78077; CAA54987.1;  
DR HSSP; P98066; ITSG.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_v.  
DR InterPro; IPR000538; Link.  
DR Pfam; PF00047; Ig; 1.  
DR Pfam; PF00193; Xlink; 2.  
DR ProDom; PD000918; Link; 2.  
DR SMART; SM00406; IGV; 1.  
DR SMART; SM00445; LINK; 2.  
DR PROSITE; PS01241; LINK; 2.  
KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;  
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.  
FT PROPEP 1 15  
FT CHAIN 16 354  
FT DOMAIN 54 146  
FT DOMAIN 176 253  
FT DOMAIN 274 350  
FT DISULFID 61 139  
FT DISULFID 181 252  
FT DISULFID 205 226

FT DISULFID 279 349 BY SIMILARITY.  
 FT DISULFID 304 325 BY SIMILARITY.  
 FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 354 AA; 40077 MW; DB15CBB7816C7172 CRC64;  
 Query Match 11.4%; Score 106.5; DB 1; Length 354;  
 Best Local Similarity 29.8%; Pred. No. 0.016;  
 Matches 37; Conservative 19; Mismatches 57; Indels 11; Gaps 6;  
 Qy 12 HLTVDYTCRPILEVPES-VTGFWKGDVNLPCY--DPL---QGYTVLVKW-LVQRGSD 63  
 Db 31 HQAENGPRLLVEAEQAKVFSHGRGNVTLPCKFLRDPFAFGSGTHKIRIKWTKLTSYDK 90  
 Qy 64 PVTFLRRSSGDHQ-QAKYQGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTPDG 122  
 Db 91 EVDVFEV--SMGYHKTKYGGYQGRVFLKGGSDNDASLVITDLTDLDYGRYKCEVIEGLEDD 148  
 Qy 123 NOV 126  
 Db 149 TAV 152  
 RESULT 2  
 PLX\_MOUSE  
 ID PLX\_MOUSE STANDARD; PRT; 356 AA.  
 AC Q9QUP5; Q9D1G9; Q921X7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Proteoglycan link protein precursor (Cartilage link protein) (LP).  
 GN CRT11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129/SV;  
 RA MEDLINE-20108790; PubMed-10640815;  
 RA Deak F., Mates L., Krysan K., Liu Z., Szabo P.E., Mann J.R.,  
 RA Beier D.R., Kiss I.;  
 RT "Characterization and chromosomal location of the mouse link protein  
 gene (Crt11).";  
 RL Cytogenet. Cell Genet. 87:75-79(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/c;  
 RA Otto J.M., Cs-Szabo G., Kamath R.V., Liu W., Li Y., Glant T.T.;  
 RT "Molecular analysis of the murine link protein gene: 5' flanking,  
 RT coding and 3' flanking sequence analysis; exon-intron structure;  
 RT tissue distribution; and generation of transgenic mice.";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6J; TISSUE-Embryo;  
 RX MEDLINE-21085660; PubMed-11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH  
 CC HYALURONIC ACID IN THE EXTRACELLULAR MATRIX.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to a  
 CC frameshift in position 347.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; AF137278; AAF24166.1;  
 DR EMBL; AF137275; AAF24166.1; JOINED.  
 DR EMBL; AF137276; AAF24166.1; JOINED.  
 DR EMBL; AF137277; AAF24166.1; JOINED.  
 DR EMBL; AF139572; AAF24977.1;  
 DR EMBL; AF098460; AAD12253.1;  
 DR EMBL; AK003382; BAB22872.1; ALT\_FRAME.  
 DR HSSP; P98066; ITSG.  
 DR MGD; MGI:1337006; Crt11.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003596; Ig\_v.  
 DR InterPro; IPR000538; Link.  
 DR Pfam; PF00047; Ig; 1.  
 DR Pfam; PF00193; Xlink; 2.  
 DR PRINTS; PR01265; LINKMODULE.  
 DR ProDom; PD000918; Link; 2.  
 DR SMART; SM00406; IGV; 1.  
 DR SMART; SM00445; Link; 2.  
 DR PROSITE; PS01241; Link; 2.  
 KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;  
 KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.  
 FT PROPEP 1 9  
 FT CHAIN 10 356 BY SIMILARITY.  
 FT DOMAIN 41 133 IG-LIKE V-TYPE DOMAIN.  
 FT DOMAIN 178 255 LINK 1.  
 FT DOMAIN 276 352 LINK 2.  
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .).  
 FT DISULFID 63 141 BY SIMILARITY.  
 FT DISULFID 183 254 BY SIMILARITY.  
 FT DISULFID 207 228 BY SIMILARITY.  
 FT DISULFID 281 351 BY SIMILARITY.  
 FT DISULFID 306 327 BY SIMILARITY.  
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 18 18 MISSING (IN REF. 2).  
 SQ SEQUENCE 356 AA; 40478 MW; 3FB4B8F30E28C9F2 CRC64;  
 Query Match 11.1%; Score 103.5; DB 1; Length 356;  
 Best Local Similarity 29.1%; Pred. No. 0.031;  
 Matches 37; Conservative 21; Mismatches 52; Indels 17; Gaps 7;  
 Qy 12 HLTVDYTCRPILEVPES-VTGFWKGDVNLPCY--DPL---QGYTVLVKWLVQRGSDPV 65  
 Db 33 HQAENGPRLLVEAEQAKVFSHGRGNVTLPCKFLRDPFAFGSGTHKIRIKW-TKLTS-- 89  
 Qy 66 TIFLRD-----SSGDHQ-QAKYQGRHLVSHKVPDVSQSLSTLEMDRSHYTCVWTP 119  
 Db 90 --YLREVDVFSVMGYHKTKYGGYQGRVFLKGGSDNDASLVITDLTDLDYGRYKCEVIEGL 147  
 Qy 120 PDGNQV 126  
 Db 148 EDDTAV 154  
 RESULT 3

A33\_HUMAN  
ID A33\_HUMAN STANDARD; PRT; 319 AA.  
AC Q99795;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cell surface A33 antigen precursor (Glycoprotein A33).  
GN GPA33.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE-Colon carcinoma;  
RX MEDLINE=97165045; PubMed=9012807;  
RA Heath J.K., White S.J., Johnstone C.N., Catimel B., Simpson R.J.,  
RA Moritz R.L., Tu G.-F., Ji H., Whitehead R.H., Groenen L.C.,  
RA Scott A.M., Ritter G., Cohen L., Welt S., Old L.J., Nice E.C.,  
RA Burgess A.W.,  
RT "The human A33 antigen is a transmembrane glycoprotein and a novel  
RT member of the immunoglobulin superfamily."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:469-474(1997).  
RN [2]  
RP POST-TRANSLATIONAL MODIFICATIONS.  
RX MEDLINE=97396159; PubMed=9245713;  
RA Ritter G., Cohen L.S., Nice E.C., Catimel B., Burgess A.W.,  
RA Moritz R.L., Ji H., Heath J.K., White S.J., Welt S., Old L.J.,  
RA Simpson R.J.,  
RT "Characterization of posttranslational modifications of human A33  
RT antigen, a novel palmitoylated surface glycoprotein of human  
RT gastrointestinal epithelium."  
RL Biochem. Biophys. Res. Commun. 236:682-686(1997).  
CC -1- FUNCTION: MAY PLAY A ROLE IN CELL-CELL RECOGNITION AND SIGNALING.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NORMAL GASTROINTESTINAL  
CC EPITHELIUM AND IN 95% OF COLON CANCERS.  
CC -1- PTM: N-GLYCOSYLATED, CONTAINS APPROXIMATELY 8 K OF N-LINKED  
CC CARBOHYDRATE.  
CC -1- PTM: PALMITOYLATED.  
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U79725; AAC50957.1;  
CC Genew; HGNC:4445; GPA33.  
CC MIN; 602171;  
CC InterPro; IPR003006; Ig\_MHC.  
CC InterPro; IPR003600; Ig\_Like.  
CC Pfam; PF00047; Ig; 2.  
CC SMART; SM00410; Ig\_Like; 1.  
CC SMART; SM00406; IgV; 1.  
CC Immunoglobulin domain; Lipoprotein; Palmitate; Glycoprotein;  
CC Transmembrane; Signal; Antigen.  
CC SIGNAL 1 21  
CC CHAIN 22 319  
CC DOMAIN 22 235  
CC TRANSMEM 236 256  
CC DOMAIN 257 319  
CC CYTOPLASMIC (POTENTIAL).  
CC IG-LIKE V-TYPE DOMAIN.  
CC IG-LIKE C2-TYPE DOMAIN.  
CC POLY-CYS.  
CC POTENTIAL.  
CC DISULFID 43 117  
CC FT 146 222

FT DISULFID 162 211  
FT CARBOHYD 112 112  
FT CARBOHYD 200 200  
FT CARBOHYD 223 223  
SQ SEQUENCE 319 AA; 35632 MW; 9BFC7AAF45C2408E CRC64;  
Query Match 11.0%; Score 102.5; DB 1; Length 319;  
Best Local Similarity 27.4%; Pred No. 0.033;  
Matches 37; Conservative 26; Mismatches 63; Indels 9; Gaps 6;  
QY 4 LIGLLLLGLHVTDTYGRPILEVPESVTPGPKG-DVNLPCYDPLQGYTVLVKW--LVQR 60  
DB 8 VLWTLCAVRVTVDAIS---VETPDVLRASQGSVTLPTCTYHTSTSSREGLIQWDKLLLT 64  
QY 61 GSDPTVIFLRDSSGHHQIAQYQGRHLVSHKV-PGDSVQLSLFLEMDRSHYTCVTWOT 119  
DB 65 HTRVVIV-PFSKNYIHGELYKNRVSISNAEQSDASITIDLTMDANGTYECSVLSMS 123  
QY 120 P-DGNQVVRDKITEL 133  
DB 124 DLEGNTKSRVRLVL 138  
RESULT 4  
PLK\_RAT  
ID PLK\_RAT STANDARD; PRT; 354 AA.  
AC P03994;  
DT 23-OCT-1986 (Rel. 02, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Proteoglycan link protein precursor (Cartilage link protein) (LP).  
GN CRTLL.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88198139; PubMed=2452158;  
RA Rhodes C., Doerge K., Sasaki M., Yamada Y.;  
RT "Alternative splicing generates two different mRNA species for rat  
RT link protein."  
RL J. Biol. Chem. 263:6063-6067(1988).  
RN [2]  
RP SEQUENCE OF 126-354 FROM N.A.  
RX MEDLINE=86233314; PubMed=3459153;  
RA Doerge K., Hassell J.R., Caterson B., Yamada Y.;  
RT "Link protein cDNA sequence reveals a tandemly repeated protein  
RT structure."  
RL Proc. Natl. Acad. Sci. U.S.A. 83:3761-3765(1986).  
RN [3]  
RP SEQUENCE OF 16-354.  
RX MEDLINE=86140139; PubMed=2419334;  
RA Neame P.J., Christner J.E., Baker J.R.;  
RT "The primary structure of link protein from rat chondrosarcoma  
RT proteoglycan aggregate."  
RL J. Biol. Chem. 261:3519-3535(1986).  
CC -1- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH  
CC HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M22340; AAA41535.1;  
CC DR

TISSUE=Embryonic sternal cartilage;  
MEDLINE=86233315; PubMed=3459154;  
Deak F., Kiss I., Sparks K.J., Argraves W.S., Hampikian G.,  
Goetnick P.F.;  
"Complete amino acid sequence of chicken cartilage link protein  
deduced from cDNA clones.";  
Proc. Natl. Acad. Sci. U.S.A. 83:3766-3770(1986).  
[2]

SEQUENCE FROM N.A.  
MEDLINE=87317659; PubMed=3476955;  
Kiss I., Deak F., Mestric S., Delius H., Soos J., Dekany K.,  
Argraves W.S., Sparks K.J., Goetnick P.F.;  
"Structure of the chicken link protein gene: exons correlate with the  
protein domains.";  
Proc. Natl. Acad. Sci. U.S.A. 84:6399-6403(1987).  
-!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH  
HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.  
-!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
-!- SIMILARITY: CONTAINS 2 LINK DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M35038; AAA48941.1; JOINED.  
EMBL; M35035; AAA48941.1; JOINED.  
EMBL; M35036; AAA48941.1; JOINED.  
EMBL; M35037; AAA48941.1; JOINED.  
EMBL; M13212; AAA48940.1;  
PIR; A24881; LKCH.  
PIR; A28305; A28305.  
HSSP; P98066; LTSG.  
InterPro; IPR003006; Ig MHC.  
InterPro; IPR003596; Ig V.  
InterPro; IPR000338; Link.  
Pfam; PF00047; Ig 1.  
Pfam; PF0193; Xlink 2.  
ProDom; PD000918; Link; 2.  
SMART; SM00406; IGV; 1.  
SMART; SM00445; LINK; 2.  
PROSITE; PS01341; LINK; 2.  
Glycoprotein; Connective tissue, Extracellular matrix; Cartilage;  
Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.  
PROPEP 1 9  
CHAIN 10 355 PROTEOGLYCAN LINK PROTEIN.  
DOMAIN 54 147 IG-LIKE V-TYPE DOMAIN.  
DOMAIN 177 254 LINK 1.  
DOMAIN 275 351 LINK 2.  
DISULFID 61 140 BY SIMILARITY.  
DISULFID 182 253 BY SIMILARITY.  
DISULFID 206 227 BY SIMILARITY.  
DISULFID 280 350 BY SIMILARITY.  
DISULFID 305 326 BY SIMILARITY.  
DISULFID 21 21 N-LINKED (GLCNAC...) (POTENTIAL).  
CARBOHYD 56 56 N-LINKED (GLCNAC...) (POTENTIAL).  
SEQUENCE 355 AA; 40533 MW; 2581CAE22158B60F CRC64;  
Query Match 10.7%; Score 100.5; DB 1; Length 355;  
Best Local Similarity 30.7%; Pred. No. 0.057;  
Matches 31; Conservative 14; Mismatches 45; Indels 11; Gaps 4;  
QY 35 GDVNLPC-----TYDPLOGYTOVLVKK--LVQRGSDPVTIFLRSSGSDHIQQ-AKYQGR 85  
DB 55 GNTVLCKFYHEHTSTAGSTGTHKIRVAKWTKLTSYLKEDVFEV--AMGHRKSYGKYQGR 112  
QY 86 LHVSHKYPGDVSLQLSLEMDRSHYTCVETWTPPDGNQVV 126  
DB 113 VFLRESENDASLIITIMLEDYGRYKCEVIEGLEDDTAVV 153

[illegible]

```

RESULT 6
PLK_HUMAN STANDARD; PRT; 354 AA.
ID PLK_HUMAN
AC P10915;
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR InterPro: IPR000538; Link.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF00193; Xlink; 2.
DR ProDom: PD000918; Link; 2.
DR SMART: SM00406; IGV; 1.
DR SMART: SM00445; LINK; 2.
DR PROSITE: PS01241; LINK; 2.
KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.
FT PROPEP 1 15
FT CHAIN 16 354
FT DOMAIN 54 146
FT DOMAIN 176 253
FT DOMAIN 274 350
FT DISULFID 61 139
FT DISULFID 181 252
FT DISULFID 205 226
FT DISULFID 279 349
FT DISULFID 304 325
FT CARBOHYD 21 21
FT CARBOHYD 56 56
SQ SEQUENCE 354 AA; 40165 MW; 315C96EC3AC2626A CRC64;
Query Match 10.78; Score 100; DB 1; Length 354;
Best Local Similarity 32.08; Pred. No. 0.064;
Matches 32; Conservative 14; Mismatches 44; Indels 10; Gaps 5;
OY 35 GDVNLPCPY--DPL---QGYTVLVK--LVQRGSDPTIFLRSSGDHIQ-QAKYQGR 86
Db 55 GNVLPCKFYRDPAFGSGIKIRIKTKLSDYLKEDVDFV--SMGYHKYGYQGRV 112
OY 87 HVSHKVPQGVLSQLSTLEMDRSHYCTVWTQTPDGNQV 126
Db 113 FLKGGSDSDASLVITDLTLEDYGRYKCEVIEGLEDVTVV 152
RESULT 7
PLK_BOVIN
ID PLK_BOVIN STANDARD; PRT; 354 AA.
AC P55252;
DR 01-OCT-1996 (Rel. 34, Created)
DR 01-OCT-1996 (Rel. 34, Last sequence update)
DR 01-NOV-1997 (Rel. 35, Last annotation update)
DE Proteoglycan link protein precursor (Cartilage link protein) (LP).
GN CRTL1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=96043243; PubMed=7584851;
RA Hering T.M., Kollar J., Huynh T.D., Sandell L.J.;
RT "Bovine chondrocyte link protein cDNA sequence: interspecies
RL conservation of primary structure and mRNA untranslated regions.";
CC Comp. Biochem. Physiol. 112B:197-203(1995).
CC -1- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH
CC HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

EMBL; U02292; AAC04311.1; -  
HSP; P98066; ITSG  
InterPro; IPR003006; Ig\_MHC  
InterPro; IPR003596; Ig\_V  
InterPro; IPR000538; Link  
Pfam; PF00047; Ig; 1  
Pfam; PF00193; Xlink; 2  
Pfam; PF000918; Link; 2  
SMART; SM00406; IGV; 1  
SMART; SM00445; LINK; 2  
PROSITE; PS01241; LINK; 2  
Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;  
Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.  
PROPEP 1 15  
CHAIN 16 354  
DOMAIN 54 146  
DOMAIN 176 253  
DOMAIN 274 350  
DISULFID 61 139  
DISULFID 181 252  
DISULFID 205 226  
DISULFID 279 349  
DISULFID 304 325  
SEQUENCE 354 AA; 40287 MW; 065D155378A1283C CRC64;  
Query Match 10.4%; Score 97.5; DB 1; Length 354;  
Best Local Similarity 28.2%; Pred. No. 0.11;  
Matches 35; Conservative 20; Mismatches 11; Gaps 6;  
2Y 12 HLTVDVTGRTILEVPES-VTGPKMGDYNLPCTY--DPL---QGYTVLVK--LVQSGD 63  
31 HIOAENGPRLLVEAEQAKVSRGGNTLPCFKYDFATFGSGTHKIRIKWTKLTSDYLK 90  
2Y 64 PVTIFLRDSSGDHIQ-QAKYQGRHLVSHKVPDVSLSLSTLEMDRSHYTCVETWQTPDG 122  
31 HIOAENGPRLLVEAEQAKVSRGGNTLPCFKYDFATFGSGTHKIRIKWTKLTSDYLK 90  
2Y 123 NQV 126  
31 HIOAENGPRLLVEAEQAKVSRGGNTLPCFKYDFATFGSGTHKIRIKWTKLTSDYLK 90  
2Y 149 TAV 152  
31 HIOAENGPRLLVEAEQAKVSRGGNTLPCFKYDFATFGSGTHKIRIKWTKLTSDYLK 90  
RESULT 8  
PLK\_PIG STANDARD; PRT; 354 AA.  
AC P10859;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Proteoglycan link protein precursor (Cartilage link protein) (LP).  
3N CRTLI.  
CS Sus scrofa (Pig).  
CX Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
CX Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
CX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Tissue-Laryngeal cartilage chondrocytes;  
RX MEDLINE=89293837; PubMed=2738916;  
RA Perkins S.J., Nealis A.S., Duthie J., Hardingham T.E.;  
RT "Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal domains and link protein."  
RL J. Mol. Biol. 206:737-753(1989).  
RC [-] FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH  
CC HYALURONIC ACID IN THE EXTRACELLULAR MATRIX.  
CC [-] SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC [-] SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
CC EMBL; Y00165; CAA68358.1;  
CC PIR; S04243; S04243.  
CC HSP; P98066; ITSG.  
CC InterPro; IPR003006; Ig\_MHC  
CC InterPro; IPR003596; Ig\_V  
CC InterPro; IPR000538; Link  
CC Pfam; PF00047; Ig; 1  
CC Pfam; PF00193; Xlink; 2  
CC Pfam; PF000918; Link; 2  
CC SMART; SM00406; IGV; 1  
CC SMART; SM00445; LINK; 2  
CC PROSITE; PS01241; LINK; 2  
CC Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;  
CC Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.  
CC PROPEP 1 9  
CC CHAIN 10 354  
CC DOMAIN 54 146  
CC DOMAIN 176 253  
CC DOMAIN 274 350  
CC DISULFID 61 139  
CC DISULFID 181 252  
CC DISULFID 205 226  
CC DISULFID 279 349  
CC DISULFID 304 325  
CC CARBOHYD 21 21  
CC CARBOHYD 56 56  
CC SEQUENCE 354 AA; 40260 MW; 58FFB9DE51ABCC1 CRC64;  
Query Match 10.4%; Score 97.5; DB 1; Length 354;  
Best Local Similarity 28.2%; Pred. No. 0.11;  
Matches 35; Conservative 20; Mismatches 58; Indels 11; Gaps 6;  
QY 12 HLTVDVTGRTILEVPES-VTGPKMGDYNLPCTY--DPL---QGYTVLVK--LVQSGD 63  
31 HIOAENGPRLLVEAEQAKVSRGGNTLPCFKYDFATFGSGTHKIRIKWTKLTSDYLK 90  
QY 64 PVTIFLRDSSGDHIQ-QAKYQGRHLVSHKVPDVSLSLSTLEMDRSHYTCVETWQTPDG 122  
31 HIOAENGPRLLVEAEQAKVSRGGNTLPCFKYDFATFGSGTHKIRIKWTKLTSDYLK 90  
QY 123 NQV 126  
31 HIOAENGPRLLVEAEQAKVSRGGNTLPCFKYDFATFGSGTHKIRIKWTKLTSDYLK 90  
QY 149 TAV 152  
31 HIOAENGPRLLVEAEQAKVSRGGNTLPCFKYDFATFGSGTHKIRIKWTKLTSDYLK 90  
RESULT 9  
BUTY\_BOVIN STANDARD; PRT; 526 AA.  
ID BUTY\_BOVIN  
AC P18952; O18955; O18959;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Butyrophilin precursor (BT).  
GN BTN1A1 OR BTN.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A. AND SEQUENCE OF 27-47.  
RX MEDLINE=90354441; PubMed=2387867;  
RA Jack L.J.W., Mather I.H.;  
RT "Cloning and analysis of cDNA encoding bovine butyrophilin, an apical  
RT glycoprotein expressed in mammary tissue and secreted in association  
RL with the milk-fat globule membrane during lactation."  
RN J. Biol. Chem. 265:14481-14486(1990).  
RC [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Holstein-Friesian;



OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



30-MAY-2000 (Rel. 39, Last sequence update)  
 15-JUN-2002 (Rel. 41, Last annotation update)  
 Coxsackievirus and adenovirus receptor precursor (Coxsackievirus B-  
 adenovirus receptor) (hCAR) (CVB3 binding protein).  
 GN CxADR OR CAR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97190109; PubMed=9036860;  
 RA Bergelson J.M., Cunningham J.A., Droguett G., Kurt-Jones E.,  
 RA Krithivas A., Hong J.S., Horwitz M.S., Crowell R.L., Finberg R.W.;  
 RT "Isolation of a common receptor for Coxsackie B viruses and  
 RT adenoviruses 2 and 5";  
 RL Science 275:1320-1323(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97250541; PubMed=9096397;  
 RA Tomko R.P., Xu R., Philipson L.;  
 RT "hCAR and hCAR: the human and mouse cellular receptors for subgroup C  
 RT adenoviruses and group B coxsackieviruses";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:3352-3356(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20008750; PubMed=10543405;  
 RA Bowles K.R., Gibson J., Wu J., Shaffer L.G., Towbin J.A.,  
 RA Bowles N.E.;  
 RT "Genomic organization and chromosomal localization of the human  
 RT Coxsackievirus B-adenovirus receptor gene";  
 RL Hum. Genet. 105:354-359(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX Anderson C.W., Kieleczawa J., Dunn J.J., Freimuth P.;  
 RT "Sequence and expression of CXADR, the human gene for the  
 RT coxsackievirus and adenovirus receptor";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases  
 CC -1- FUNCTION: SERVES AS A RECEPTOR FOR GROUP B COXSACKIEVIRUSES AND  
 CC SUBGROUP C OF ADENOVIRUSES (AD2 AND AD5).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use, by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; Y07593; CAA68868.1; -  
 DR EMBL; U0716; AAC51234.1; -  
 DR EMBL; AF169366; AAF05908.1; -  
 DR EMBL; AF169360; AAF05908.1; JOINED.  
 DR EMBL; AF169361; AAF05908.1; JOINED.  
 DR EMBL; AF169362; AAF05908.1; JOINED.  
 DR EMBL; AF169363; AAF05908.1; JOINED.  
 DR EMBL; AF169364; AAF05908.1; JOINED.  
 DR EMBL; AF169365; AAF05908.1; JOINED.  
 DR EMBL; AF200465; AAF24344.1; -  
 DR Genew; HGNC:2559; CXADR.  
 DR MIM; 602621; -  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003598; Ig\_C2.  
 DR InterPro; IPR003600; Ig\_Like.  
 DR Pfam; PF00047; Ig; 2.  
 DR SMART; SM00410; IG\_Like; 1.  
 DR SMART; SM00408; IG2; 1.  
 KW Immunoglobulin domain; Receptor; Transmembrane; Glycoprotein; Signal;  
 KW Repeat.  
 FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 365  
 FT DOMAIN 20 237  
 FT TRANSMEM 238 258  
 FT TRANSMEM 259 365  
 FT DOMAIN 34 127  
 FT DOMAIN 155 219  
 FT DISULFID 41 120  
 FT DISULFID 162 212  
 FT CARBOHYD 106 106  
 FT CARBOHYD 201 201  
 SQ SEQUENCE 365 AA; 40029 MW; AB01C6346CB7FE64 CRC64;  
 Query Match 10.28; Score 95.5; DB 1; Length 365;  
 Best Local Similarity 26.08; Pred. No. 0.17;  
 Matches 53; Conservative 28; Mismatches 80; Indels 43; Gaps 13;  
 QY 1 MGILLGLLLGLHFLVDYGRPI-LEVPESVTGPMKGD-VNLPCTV-----DPLQGYTQ 51  
 DB 1 MALLLCFVLL--CGVDFARSLSTTPEEMIEKAGETAYLPCKFTLSPEDQGPLD---- 54  
 QY 52 VLVKWLVRGSG---DPVTIFLRDSSGDHIQQAQY---QGRHVSFK--VPGDVSLSLST 102  
 DB 55 --IEWLSPADNQKVDQVILY---SGDKYDYYDPLKGRVHFTSNDLKSGLDASINVTN 109  
 QY 103 LEMDDRSHTCEVTWQTPDGNQ---VVRKITELRVQKHSSKLL---KTKTEAPTMT 154  
 DB 110 LQSLSDIGTYQCKVKKAPGVANKIHLVLPKSPGARGCYVDGSEEGISDFKIKCP-PREGS 168  
 QY 155 YPL-----KATSTVKQSWDMWTTDM 173  
 DB 169 LPLOYEWQKLSDSOKMPTSLAEM 192  
 RESULT 13  
 MOG\_RAT-  
 ID MOG\_RAT STANDARD; PRT; 245 AA.  
 AC Q63345;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myelin-oligodendrocyte glycoprotein precursor.  
 GN MOG.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RX MEDLINE=93085763; PubMed=1453482;  
 RA Gardiner M.V., Amiguet P., Linington C., Matthieu J.-M.;  
 RT "Myelin/oligodendrocyte glycoprotein is a unique member of the  
 RT immunoglobulin superfamily";  
 RL J. Neurosci. Res. 33:177-187(1992).  
 RN [2]  
 RP SEQUENCE OF 28-245 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93376728; PubMed=8367453;  
 RA Pham-Dinh D., Mattei M.-G., Nussbaum J.-L., Roussel G.,  
 RA Pontarotti P., Roedel N., Mather I.H., Artzt K., Lindahl K.F.,  
 RA Dautigny A.;  
 RT "Myelin/oligodendrocyte glycoprotein is a member of a subset of the  
 RT immunoglobulin superfamily encoded within the major  
 RT histocompatibility complex";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7990-7994(1993).  
 RN [3]  
 RP STRUCTURE BY NMR OF 62-82.  
 RX MEDLINE=97354172; PubMed=9210466;  
 RA Albouze Abo S., Wilson J.C., Bernard C.C.A., von Itzstein M.;  
 RT "A conformational study of the human and rat encephalitogenic myelin  
 RT oligodendrocyte glycoprotein peptides 35-55";  
 RL Eur. J. Biochem. 246:59-70(1997).  
 CC -1- FUNCTION: MINOR COMPONENT OF THE MYELIN SHEATH. MAY BE INVOLVED IN  
 CC COMPLETION AND/OR MAINTENANCE OF THE MYELIN SHEATH AND IN CELL-

CC CELL COMMUNICATION.  
 CC -1- SUBUNIT: MAY FORM HOMODIMERS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE CNS, WHERE IT IS  
 CC LOCALIZED ON THE SURFACE OF MYELIN AND OLIGODENDROCYTE CYTOPLASMIC  
 CC MEMBRANES.  
 CC -1- DEVELOPMENTAL STAGE: A PEAK OF EXPRESSION HAS BEEN OBSERVED  
 CC BETWEEN POSTNATAL DAYS 15 AND 25, COINCIDENT WITH THE PERIOD OF  
 CC ACTIVE MYELINATION.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. BTN/MOG  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC -1- CAUTION: DO NOT CONFUSE MYELIN-OLIGODENDROCYTE GLYCOPROTEIN (MOG)  
 CC WITH OLIGODENDROCYTE-MYELIN GLYCOPROTEIN (OMG).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M99485; AAA41628.1; -  
 CC EMBL: L21995; AAF74786.1; -  
 CC InterPro: IPR003006; Ig\_MHC.  
 CC Pfam: PF00047; Ig\_1.  
 CC SMART: SM00406; Igv; 1.  
 CC Immunoglobulin domain; Glycoprotein; Myelin; Transmembrane; Signal.  
 KW SIGNAL 1 27  
 FT CHAIN 28 245 MYELIN-OLIGODENDROCYTE GLYCOPROTEIN.  
 FT DOMAIN 28 155 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 156 176 POTENTIAL.  
 FT DOMAIN 177 208 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 209 229 POTENTIAL.  
 FT DOMAIN 230 245 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 44 132 IG-LIKE V-TYPE DOMAIN.  
 FT DISULFID 51 125 POTENTIAL.  
 FT CARBOHYD 58 58 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 245 AA; 27881 MW; C97F8AD60D6A32B4 CRC64;  
 Query Match 9.9%; Score 92.5; DB 1; Length 245;  
 Best Local Similarity 27.5%; Pred. No. 0.2;  
 Matches 33; Conservative 16; Mismatches 54; Indels 17; Gaps 4;  
 QY 4 LGLGLL-----GHLVDYGRPILEVPSVTGPWGDVNLCTYDPLCYTOVLVK 55  
 DB 13 LLSLLLLQLLSRYAGQFRVIGPHI---RALVG---DEALPGRISPGKNATGMEVG 65  
 QY 56 WLVRGSDPVTIFLRDSSGDHIQQAQYGRHLVSHKVPGD--VSLQSLTLEMDRSHYTC 113  
 DB 66 WYRSPFRVHLRYNGKDDAQEAPEYRGRTLLKESIGKVALRIQNVRSEGGYTC 125  
 RESULT 14  
 PGCV\_BOVIN STANDARD; PRT: 3381 AA.  
 ID PGCV\_BOVIN 077609; 077610; 077611; 077612;  
 AC P81282; 077609; 077610; 077611; 077612;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Versican core protein precursor (Large fibroblast proteoglycan)  
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial  
 DE hyaluronate-binding protein) (GHAP).  
 GN CP5G2.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS VO; V1; V2 AND V3).

RC TISSUE-Forebrain;  
 RX MEDLINE=98288320; PubMed=9624174;  
 RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,  
 RT Zimmermann D.R.;  
 RA "Versican V2 is a major extracellular matrix component of the mature  
 RT bovine brain.";  
 RL J. Biol. Chem. 273:15758-15764(1998).  
 RN [2]  
 RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331  
 RP AND 342-348.  
 RC TISSUE-Spinal cord;  
 RX MEDLINE=92062692; PubMed=1720020;  
 RA Perides G., Biviano F., Bignami A.;  
 RT "Interaction of a brain extracellular matrix protein with hyaluronic  
 RT acid.";  
 RL Biochim. Biophys. Acta 1075:248-258(1991).  
 CC -1- FUNCTION: May play a role in intercellular signaling and in  
 CC connecting cells with the extracellular matrix. May take part in  
 CC the regulation of cell motility, growth and differentiation. Binds  
 CC hyaluronic acid.  
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2  
 CC and V3; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed  
 CC in the central nervous system, and in a number of mesenchymal and  
 CC epithelial tissues; the major isoform V2 is restricted to the  
 CC central nervous system.  
 CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development  
 CC (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF060456; AAC24358.1; -  
 CC EMBL: AF060457; AAC24359.1; -  
 CC EMBL: AF060458; AAC24360.1; -  
 CC EMBL: AF060459; AAC24361.1; -  
 CC HSP: P01132; 1EPG.  
 CC InterPro: IPR00152; Asx\_hydroxyl.  
 CC InterPro: IPR000561; EGF-like.  
 CC InterPro: IPR000742; EGF\_2.  
 CC InterPro: IPR001881; EGF\_Ca.  
 CC InterPro: IPR003599; Ig.  
 CC InterPro: IPR003006; Ig\_MHC.  
 CC InterPro: IPR003600; Ig\_Like.  
 CC InterPro: IPR001304; Lectin\_C.  
 CC InterPro: IPR000538; Link.  
 CC InterPro: IPR000436; Sushi\_SCR\_CCP.  
 CC Pfam: PF00008; EGF; 2.  
 CC Pfam: PF00047; Ig; 1.  
 CC Pfam: PF00059; lectin\_c; 1.  
 CC Pfam: PF00084; sushi; 1.  
 CC Pfam: PF00193; xlink; 2.  
 CC PRINTS: PR01265; LINKMODULE.  
 CC ProDom: PD000918; Link; 1.  
 CC SMART: SM00032; CCP; 1.  
 CC SMART: SM00034; CLECT; 1.  
 CC SMART: SM00181; EGF; 2.  
 CC SMART: SM00179; EGF\_CA; 2.  
 CC SMART: SM00409; IG; 1.  
 CC SMART: SM00410; IG\_Like; 1.  
 CC SMART: SM00445; LINK; 2.

DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
DR PROSITE; PS0041; C-TYPE LECTIN\_2; 1.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 1.  
DR PROSITE; PS01241; LINK; 2.  
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
KW Hyaluronic acid; Alternative splicing.  
FT SIGNAL 1 20  
FT CHAIN 21 3381  
FT DOMAIN 37 138  
FT DOMAIN 168 245  
FT DOMAIN 266 347  
FT DOMAIN 349 1336  
FT DOMAIN 1337 3074  
FT DOMAIN 3074 3110  
FT DOMAIN 3112 3148  
FT DOMAIN 3161 3275  
FT DOMAIN 3280 3338  
FT DISULFID 44 131  
FT DISULFID 173 244  
FT DISULFID 197 218  
FT DISULFID 271 346  
FT DISULFID 295 316  
FT DISULFID 3078 3089  
FT DISULFID 3083 3098  
FT DISULFID 3100 3109  
FT DISULFID 3116 3127  
FT DISULFID 3121 3136  
FT DISULFID 3138 3147  
FT DISULFID 3154 3165  
FT DISULFID 3182 3274  
FT DISULFID 3250 3266  
FT DISULFID 3281 3324  
FT DISULFID 3310 3337  
FT CARBOHYD 57 57  
FT CARBOHYD 331 331  
FT CARBOHYD 352 352  
FT CARBOHYD 817 817  
FT CARBOHYD 965 965  
FT CARBOHYD 1017 1017  
FT CARBOHYD 1333 1333  
FT CARBOHYD 1393 1393  
FT CARBOHYD 1437 1437  
FT CARBOHYD 1463 1463  
FT CARBOHYD 1653 1653  
FT CARBOHYD 1874 1874  
FT CARBOHYD 2045 2045  
FT CARBOHYD 2074 2074  
FT CARBOHYD 2103 2103  
FT CARBOHYD 2263 2263  
FT CARBOHYD 2290 2290  
FT CARBOHYD 2356 2356  
FT CARBOHYD 2623 2623  
FT CARBOHYD 2841 2841  
FT CARBOHYD 2919 2919  
FT CARBOHYD 3052 3052  
FT CARBOHYD 3354 3354  
FT CARBOHYD 3364 3364  
FT VARSPPLIC 349 349  
FT VARSPPLIC 350 1336  
FT VARSPPLIC 1337 3074  
FT VARSPPLIC 350 3074  
FT CONFLICT 25 25  
FT CONFLICT 51 51  
FT CONFLICT 89 89  
FT CONFLICT 96 96  
FT CONFLICT 346 346  
SQ SEQUENCE 3381 AA; 369984 MW; F0916FA7778D459 CRC64;

Query Match 9.9%; Score 92.5; DB 1; Length 3381;  
Best Local Similarity 30.0%; Pred No. 4.9;  
Matches 36; Conservative 14; Mismatches 53; Indels 17; Gaps 7;  
QY 23 LEVPESTVTPGPKGVDNLPFC-----TYDPLQGYTQ--VLVKW----LVQRGSD--PVTF 68  
Db 26 MEKSPVPKGLSGKYNLPCHFTMTPLPSYNTTSEFLRIKWSKIELDKTGKDLKETTIVL 85  
QY 69 LRDSSGDHIQQAQYGRHLV-SH-KVPGDVSQLSLTLEMDRSHYTCVETWTPPGNQVV 126  
Db 86 VAQNGNIKIGQ-DYKGRVSVTPHPEDVDGASLTMVKLLASDAGRYRCVDMYGIEDTQDTV 144

RESULT 15  
CD86\_MOUSE  
ID CD86\_MOUSE STANDARD; PRT: 309 AA.  
AC P42082;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE T lymphocyte activation antigen CD86 precursor (Activation B7-2  
DE antigen) (Early T cell costimulatory molecule-1) (ETC-1).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94065585; PubMed=7504059;  
RA Freeman G.J., Borriello F., Hodes R.J., Reiser H., Gribben J.G.,  
RA Ng J.W., Kim J., Goldberg J.M., Hathcock K., Laszlo G., Lombard L.A.,  
RA Wang S., Gray G.S., Nadler L.M., Sharpe A.H.;  
RT "Murine B7-2, an alternative CTLA4 counter-receptor that costimulates  
RT T cell proliferation and interleukin 2 production.";  
RL J. Exp. Med. 178:2185-2192(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96094437; PubMed=7499829;  
RA Borriello F., Oliveros J., Freeman G.J., Nadler L.M., Sharpe A.H.;  
RT "Differential expression of alternate mB7-2 transcripts.";  
RL J. Immunol. 155:5490-5497(1995).  
RN [3]  
RP SEQUENCE OF 7-309 FROM N.A.  
RX MEDLINE=94230971; PubMed=7513726;  
RA Chen C., Gault A., Shen L., Nabavi N.;  
RT "Molecular cloning and expression of early T cell costimulatory  
RT molecule-1 and its characterization as B7-2 molecule.";  
RL J. Immunol. 152:4929-4936(1994).  
CC -!- FUNCTION: RECEPTOR INVOLVED IN THE COSTIMULATORY SIGNAL ESSENTIAL  
CC FOR T LYMPHOCYTE PROLIFERATION AND INTERLEUKIN 2 PRODUCTION, BY  
CC BINDING CD28 OR CTLA-4. MAY PLAY A CRITICAL ROLE IN THE EARLY  
CC EVENTS OF T CELL ACTIVATION AND COSTIMULATION OF NAIVE T CELLS,  
CC SUCH AS DECIDING BETWEEN IMMUNITY AND ANERGY THAT IS MADE BY T  
CC CELLS WITHIN 24 HOURS AFTER ACTIVATION.  
CC -!- SUBCELLULAR LOCATION: type I membrane protein.  
CC -!- TISSUE SPECIFICITY: EXPRESSED ON ACTIVATED B CELLS.  
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-7 IS THE INITIATOR.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: L25606; AAA79770.1;  
DR EMBL: U39456; AAC52334.1;

Search completed: March 4, 2003, 15:00:00.  
Job time : 5.18902 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:32 ; Search time 11.9893 Seconds  
(without alignments)  
3007.543 Million cell updates/sec

Title: US-09-763-902b-6

Perfect score: 936

Sequence: 1 MGILLGLLLGLHGLTVDTYGR.....PLKATSTVKQSWDWTMDMG 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 870   | 92.9        | 399    | 4 Q9Y279  | Q9Y279 homo sapien |
| 2          | 130.5 | 13.9        | 360    | 4 Q96S86  | Q96S86 homo sapien |
| 3          | 109.5 | 11.7        | 319    | 6 Q9TU80  | Q9TU80 canis famil |
| 4          | 109   | 11.6        | 390    | 4 Q9HIX9  | Q9HIX9 homo sapien |
| 5          | 109   | 11.6        | 512    | 4 Q96DN8  | Q96DN8 homo sapien |
| 6          | 109   | 11.6        | 536    | 4 Q96RW7  | Q96RW7 homo sapien |
| 7          | 107   | 11.4        | 412    | 4 Q96PW2  | Q96PW2 homo sapien |
| 8          | 104.5 | 11.2        | 272    | 11 Q70356 | Q70356 mus musculu |
| 9          | 104.5 | 11.2        | 365    | 6 Q8WVY3  | Q8WVY3 bos taurus  |
| 10         | 102   | 10.9        | 549    | 4 Q9NQS3  | Q9NQS3 homo sapien |
| 11         | 101   | 10.8        | 335    | 13 Q9YGH1 | Q9YGH1 gallus gall |
| 12         | 101   | 10.8        | 335    | 13 Q9PWR4 | Q9PWR4 gallus gall |
| 13         | 100.5 | 10.7        | 282    | 11 Q8VIM1 | Q8VIM1 mus musculu |
| 14         | 100   | 10.7        | 305    | 11 Q8VIM2 | Q8VIM2 mus musculu |
| 15         | 100   | 10.7        | 335    | 13 Q9YGV5 | Q9YGV5 gallus gall |
| 16         | 100   | 10.7        | 438    | 11 Q9JLB7 | Q9JLB7 mus musculu |

|    |      |      |      |           |                    |
|----|------|------|------|-----------|--------------------|
| 17 | 100  | 10.7 | 510  | 11 Q9JLB8 | Q9JLB8 mus musculu |
| 18 | 100  | 10.7 | 549  | 11 Q9JLB9 | Q9JLB9 mus musculu |
| 19 | 97.5 | 10.4 | 286  | 6 Q46535  | Q46535 bos taurus  |
| 20 | 97.5 | 10.4 | 453  | 16 Q9IIV4 | Q9IIV4 pseudomonas |
| 21 | 96.5 | 10.3 | 549  | 11 Q9D006 | Q9D006 mus musculu |
| 22 | 96   | 10.3 | 526  | 4 Q9H458  | Q9H458 homo sapien |
| 23 | 95.5 | 10.2 | 344  | 4 Q9UKV4  | Q9UKV4 homo sapien |
| 24 | 95   | 10.1 | 307  | 11 Q54947 | Q54947 rattus norv |
| 25 | 95   | 10.1 | 731  | 6 Q8SPI6  | Q8SPI6 macropus eu |
| 26 | 94.5 | 10.1 | 181  | 11 Q9D7V7 | Q9D7V7 mus musculu |
| 27 | 94   | 10.0 | 164  | 11 Q99KG0 | Q99KG0 mus musculu |
| 28 | 93.5 | 10.0 | 318  | 13 Q91664 | Q91664 xenopus lae |
| 29 | 93.5 | 10.0 | 319  | 6 Q9TU79  | Q9TU79 sus scrofa  |
| 30 | 93.5 | 10.0 | 439  | 13 Q57349 | Q57349 gallus gall |
| 31 | 93   | 9.9  | 313  | 11 Q35531 | Q35531 rattus norv |
| 32 | 92   | 9.8  | 314  | 11 Q61238 | Q61238 mus musculu |
| 33 | 92   | 9.8  | 356  | 11 Q64381 | Q64381 mus musculu |
| 34 | 92   | 9.8  | 373  | 4 Q9H6B4  | Q9H6B4 homo sapien |
| 35 | 91   | 9.7  | 235  | 11 Q8VH42 | Q8VH42 sigmodon hi |
| 36 | 91   | 9.7  | 531  | 11 Q62319 | Q62319 mus musculu |
| 37 | 90.5 | 9.7  | 200  | 4 Q8WWT7  | Q8WWT7 homo sapien |
| 38 | 90.5 | 9.7  | 252  | 4 Q8WWT6  | Q8WWT6 homo sapien |
| 39 | 90   | 9.6  | 243  | 4 Q8TAG5  | Q8TAG5 homo sapien |
| 40 | 90   | 9.6  | 351  | 5 Q97111  | Q97111 drosophila  |
| 41 | 90   | 9.6  | 449  | 4 Q9UEI6  | Q9UEI6 homo sapien |
| 42 | 90   | 9.6  | 1072 | 2 Q86478  | Q86478 streptococ  |
| 43 | 89.5 | 9.6  | 305  | 4 Q96A60  | Q96A60 homo sapien |
| 44 | 89.5 | 9.6  | 309  | 11 Q91YV7 | Q91YV7 mus musculu |
| 45 | 88.5 | 9.5  | 235  | 11 Q8R0A6 | Q8R0A6 mus musculu |

## ALIGNMENTS

### RESULT 1

Q9Y279 PRELIMINARY; PRT: 399 AA.  
AC Q9Y279;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE 239IG protein precursor.  
GN 239IG OR DA159A1.1  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Langnese K., Kloos D.U., Colleaux L., Fontes M., Wleackner P.;  
RT "Cloning of a novel human Ig superfamily protein.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RA Pearce A.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ132502; CAB51536.1;  
DR EMBL; AL034397; CAB46921.1;  
DR InterPro; IPR003598; IG\_C2.  
DR InterPro; IPR003598; IG\_C2.  
DR InterPro; IPR003598; IG\_C2.  
DR InterPro; IPR003598; IG\_C2.  
DR Pfam; PF00047; Ig; 2.  
DR SMART; SM00408; IGC2; 1.  
DR SMART; SM00410; IG-like; 1.  
KW Immunoglobulin domain; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 399  
SQ SEQUENCE 399 AA; 43987 MW; 735CA3BC58185035 CRC64;  
Z39IG PROTEIN.

Query Match 92.9%; Score 870; DB 4; Length 399;  
Best Local Similarity 64.7%; Pred. No. 6.6e-77;  
Matches 174; Conservative 0; Mismatches 1; Indels 94; Gaps 1;

```

QY 1 MGILLGALLLGHLLTVDTYGRPILEVPESTGPKGDNVLPCTYDPLQGYTVLVKWLVR 60
DE ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MGILLGALLLGHLLTVDTYGRPILEVPESTGPKGDNVLPCTYDPLQGYTVLVKWLVR 60
QY 61 GSDPTIFLRDSSGHHIQAKYQGRHLVSHKVPDVSQLSLSTLEMDRSHYTCVWTQTP 120
DE ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 GSDPTIFLRDSSGHHIQAKYQGRHLVSHKVPDVSQLSLSTLEMDRSHYTCVWTQTP 120
QY 121 DGNQVVRDKITELRVQ----- 136
DE ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 DGNQVVRDKITELRVQKLSVSPVTYTGSGYFTVPQGMRISSLOCAQSGPPISTIWKYQ 180
QY 137 -----KHSSKLLKTK 146
DE ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 QTNNOEPIKVAITLLFPKPAVIADSGSVFCTAKQGVSEHQSHSDIVKFPVKDSSKLLKTK 240
QY 147 TEAPTMTYPLKATSTVKQSWDWTMDG 175
DE ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 TEAPTMTYPLKATSTVKQSWDWTMDG 269

RESULT 2
ID Q96S86 PRELIMINARY; PRT; 360 AA.
AC Q96S86;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Proteoglycan link protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li N., Zhang W., Wan T., Zhang M., Cao X.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037161; AAK67639.1;
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR00538; Link.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF0193; Xlink; 2.
DR ProDom; PD000918; Link; 2.
DR PROSITE; PS01241; LINK; UNKNOWNL.2.
SQ SEQUENCE 360 AA; 40894 MW; 3B72AE88D0D9E8DC CRC64;

Query Match 13.9%; Score 130.5; DB 4; Length 360;
Best Local Similarity 28.4%; Pred. No. 0.00015;
Matches 52; Conservative 20; Mismatches 42; Indels 69; Gaps 10;

QY 1 MGILLGALLLGHLLTVDTYGRP-----ILEVPESVT 30
DB 1 MGILLLVPLL--LLPGSYGLPYNGFYNSANDQNLGNHGCKDLNGVKLVETPEETL 58
QY 31 GPWKQ-DVNLPQ--TYDP-LOGYTOVLVKW--LVQSGDP-----VTIFLRDSS-GDHIQ 79
DB 59 FYQAGSVILPCRYEYALVSPRVKWKWLSENGAPEKDVILVAILGRHSFGD---- 114
QY 80 AKYQGRHLVSHKVPDVSQLSLSTLEMDRSHYTCV-----TWQ 119
DB 115 --YGRVHLRQDKHEHVSLEIQDLRELDYGRVCEVIDGLESDGLVELELRGVVFPQS 172
QY 120 PDG 122
DB 173 PNG 175

RESULT 3
Q9TU80 PRELIMINARY; PRT; 319 AA.
ID Q9TU80;
AC Q9TU80;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

```

```

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Coxsackie-adenovirus-receptor homolog (Fragment).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRED GERMAN SHEPHERD; TISSUE=LIVER;
RX MEDLINE=99422053; PubMed=10490761;
RA Fechner H., Haack A., Wang H., Wang X., Eizema K., Pauschinger M.,
RA Schoemaker R.G., van Veghel R., Houtsmuller A.B., Schultheiss H.-P.,
RA Lamers J.M.J., Poller W.;
RT "Expression of Coxsackie-adenovirus receptor and alpha v-integrin does
RT not correlate with adenovector targeting in vivo indicating anatomical
RT vector barriers."
RL Gene Ther. 6:1520-1535(1999).
DR EMBL; AF109645; AAF01256.1;
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003600; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00410; IG-like; 1.
KW Immunoglobulin domain; Receptor.
FT NON_TER 319 319
SQ SEQUENCE 319 AA; 35422 MW; 85C63A6EC7986965 CRC64;

Query Match 11.7%; Score 109.5; DB 6; Length 319;
Best Local Similarity 27.5%; Pred. No. 0.015;
Matches 56; Conservative 28; Mismatches 77; Indels 43; Gaps 13;

QY 1 MGILLGALLLGHLLTVDTYGRPI-LEVPESTGPKWGD-VNLPCTY-----DPLQGYQ 51
DB 1 MALLRFVLL--CGVADFTRSLITPEQMIEKAKETAYLPCKFTLSPEDQGPLD---- 54
QY 52 VLVKVLVQGS----DPVTIFLRDSSGDHIQAKYQ---GRLHVS HK--VPGDVSQLSLST 102
DB 55 --IEWLLSPADNOKVDQVILY---SGDKIYDDYQDLKGRVHTSNDLKSGDASINVTN 109
QY 103 LEMDRSHYTCVWTQTPDGNQ-----VVRDKITELRVQKHSHSKLL-----KTKTEAPTMT 154
DB 110 LRLSDIGTYQCKKAPGVGNKKIQLTVLVRPSGRCYVDGSEEGNDFKLKCE-PKES 168
QY 155 YPL-----KATSTVKQSWDWTDM 173
DB 169 LPQYEWQKLSNSQKTPPSWSTDM 192

RESULT 4
Q9H1X9 PRELIMINARY; PRT; 390 AA.
ID Q9H1X9;
AC Q9H1X9;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE BG15303.1 (Similar to C.elegans hemicentin) (Fragment).
GN BG15303.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lawlor S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133515; CAC17584.1;
DR HSSP; P56276; ITK.
DR InterPro; IPR003598; Ig.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003600; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 5.

```

DR SMART: SM00409; IG: 4.  
DR SMART: SM00408; IGC2: 4.  
DR SMART: SM00410; IG: like: 1.  
KW Immunoglobulin domain.  
FT NON\_TER 390 390  
SQ SEQUENCE 390 AA: 42153 MW: 51785DSBEGE19B4 CRC64;  
Query Match 11.6%; Score 109; DB 4; Length 390;  
Best Local Similarity 19.6%; Pred. No. 0.022;  
Matches 53; Conservative 33; Mismatches 73; Indels 112; Gaps 9;  
QY 4 LLGLLL-----GHLTVDTYGRPI-LEVPESVTGPKMGDYNL 39  
Db 127 LLGLLKIQETQDLADGYTCVAINAGRATGKITLDVGSPPVFIQEPADVSMEIGSNVTL 186  
QY 40 PCYDPLQGYTVLVKLVKLVQSGDPVTFILRDS-----GDHI 77  
Db 187 PCY---VQGYPEPTIKW---RRLDNMFIKSPFVSISQLRTGALFILNLWASDKGTI 240  
QY 78 QOAKYQ-GRLLH-----VSHKVPG-----94  
Db 241 CEANQFGKIOSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPCTLLAGNPPIERRWIK 300  
QY 95 -----DVSLSLTLEMDRSHYTCVETWQTPDGNQVVRDKITELRVOKHS 139  
Db 301 NSAMLLQNPIYTVRSQSLHIERVOLQDGGEYTCVSNVAGTNNKTTSSVVHVLPITQHG 360  
QY 140 SKLLTKTEAPTMTYPLKATSTVTKQSDWT 170  
Db 361 QOILSTIEGIPVTL--PCKASGNPKPSVIWS 389  
RESULT 5  
Q96DN8 PRELIMINARY; PRT; 512 AA.  
AC Q96DN8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE CNA FLJ31774 fis, clone NT2R1208030, weakly similar to basement  
DE membrane-specific heparan sulfate proteoglycan core protein  
DE precursor.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,  
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
RA Wagatsuma M., Takahashi-Fujii A., Oshina A., Sugiyama A., Kawakami B.,  
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isoqal T.;  
RA "NEDO human cDNA sequencing project."  
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AK056336; BAB71154.1;  
DR InterPro: IPR003006; IG\_MHC.  
DR InterPro: IPR003596; IG\_V.  
DR Pfam: PF00047; IG: 6.  
DR SMART: SM00406; IG: 2.  
SQ SEQUENCE 512 AA: 54971 MW: 9368150E8A5BD6C8 CRC64;  
Query Match 11.6%; Score 109; DB 4; Length 512;  
Best Local Similarity 19.6%; Pred. No. 0.031;  
Matches 53; Conservative 33; Mismatches 73; Indels 112; Gaps 9;  
QY 4 LLGLLL-----GHLTVDTYGRPI-LEVPESVTGPKMGDYNL 39  
Db 94 LLGLLKIQETQDLADGYTCVAINAGRATGKITLDVGSPPVFIQEPADVSMEIGSNVTL 153

QY 40 PCTYDPLQGYTVLVKLVKLVQSGDPVTFILRDS-----GDHI 77  
Db 154 PCY---VQGYPEPTIKW---RRLDNMFIKSPFVSISQLRTGALFILNLWASDKGTI 207  
QY 78 QOAKYQ-GRLLH-----VSHKVPG-----94  
Db 208 CEANQFGKIOSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPCTLLAGNPPIERRWIK 267  
QY 95 -----DVSLSLTLEMDRSHYTCVETWQTPDGNQVVRDKITELRVOKHS 139  
Db 268 NSAMLLQNPIYTVRSQSLHIERVOLQDGGEYTCVSNVAGTNNKTTSSVVHVLPITQHG 327  
QY 140 SKLLTKTEAPTMTYPLKATSTVTKQSDWT 170  
Db 328 QOILSTIEGIPVTL--PCKASGNPKPSVIWS 356  
RESULT 6  
Q96RW7 PRELIMINARY; PRT; 5636 AA.  
AC Q96RW7;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hemocentin.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Trent J.;  
RT "Human hemocentin gene";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF156100; AAK68690.1;  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000875; Cecropin.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR003006; IG\_MHC.  
DR InterPro: IPR000169; SHprot\_acsite.  
DR InterPro: IPR000884; TSPL.  
DR Pfam: PF00008; EGF; 5.  
DR Pfam: PF00047; IG: 44.  
DR Pfam: PF00090; tsp\_1; 6.  
DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_5.  
DR PROSITE: PS00268; CECROPIN; UNKNOWN\_1.  
DR PROSITE: PS01186; EGF\_2; UNKNOWN\_3.  
DR PROSITE: PS01187; EGF\_CA; UNKNOWN\_8.  
DR PROSITE: PS00639; THIOL\_PROTEASE\_HIS; UNKNOWN\_1.  
DR PROSITE: PS00092; TSPL; 6.  
SQ SEQUENCE 5636 AA: 613660 MW: F000B319CED7B52C CRC64;  
Query Match 11.6%; Score 109; DB 4; Length 5636;  
Best Local Similarity 19.6%; Pred. No. 0.75;  
Matches 53; Conservative 33; Mismatches 73; Indels 112; Gaps 9;  
QY 4 LLGLLL-----GHLTVDTYGRPI-LEVPESVTGPKMGDYNL 39  
Db 753 LLGLLKIQETQDLADGYTCVAINAGRATGKITLDVGSPPVFIQEPADVSMEIGSNVTL 812  
QY 40 PCTYDPLQGYTVLVKLVKLVQSGDPVTFILRDS-----GDHI 77  
Db 813 PCY---VQGYPEPTIKW---RRLDNMFIKSPFVSISQLRTGALFILNLWASDKGTI 866  
QY 78 QOAKYQ-GRLLH-----VSHKVPG-----94  
Db 867 CEANQFGKIOSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPCTLLAGNPPIERRWIK 926  
QY 95 -----DVSLSLTLEMDRSHYTCVETWQTPDGNQVVRDKITELRVOKHS 139  
Db 927 NSAMLLQNPIYTVRSQSLHIERVOLQDGGEYTCVSNVAGTNNKTTSSVVHVLPITQHG 986

```
QY 140 SKLKTKEAPTMTYPLKATSTVKQSDWT 170
DB 987 QQILSTEGIPVTL--PCKASGNPKPSVINS 1015

RESULT 7
Q96PW2 PRELIMINARY; PRT; 412 AA.
ID Q96PW2
AC Q96PW2
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE KIAA1926 protein (Fragment).
GN KIAA1926.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RT Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.
RL DNA Res. 8:179-187(2001).
DR EMBL: AB067513; BAB67819.1;
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR000538; Link.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF00193; Xlink; 2.
DR ProDom: PD000918; Link; 2.
DR PROSITE: PS01241; Link; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 412 AA; 43652 MW; 3657E0AB81BBCA73 CRC64;

Query Match 11.4%; Score 107; DB 4; Length 412;
Best Local Similarity 28.5%; Pred. No. 0.037;
Matches 35; Conservative 15; Mismatches 57; Indels 16; Gaps 5;

QY 11 GHLTVDTYGRPILEVPSVTPGPKGVNLCCTVD---PLOGYQVLYKWLVRQSDPVT- 66
DB 55 GSVVQVT-----APGVVSHRGTTVLCRYHYEAAGHGDLKWLK--TKVVDPLAF 105
QY 67 --IFLRSSGDHIQQAQYQGRHLVSHRVPDGVSLQSLTLEMDRSHVTCVTTWQTPDGNQ 124
DB 106 TDVFAV-LGQHRAFSGYRGAELOGDGPASLVLNRVTLQDYGRYCEVTNELEDAG 164
QY 125 VVR 127
DB 165 MVK 167

RESULT 8
O70356 PRELIMINARY; PRT; 272 AA.
ID O70356
AC O70356
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Butyrophilin-like (Fragment).
GN BTNL2 OR NG10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Qin S., Loretz C., Mix L., Lasky S., Madan A., Hood L.;
RT "Sequence of the mouse major histocompatibility class II region."
RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL: AF050157; AAC05289.1;
DR MGD; MGI:1859549; Btl12.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; Ig; 1.
FT NON_TER 272
SQ SEQUENCE 272 AA; 30335 MW; E745E84220343663 CRC64;

Query Match 11.2%; Score 104.5; DB 11; Length 272;
Best Local Similarity 27.1%; Pred. No. 0.038;
Matches 39; Conservative 23; Mismatches 61; Indels 21; Gaps 6;

QY 2 GILLGLLLL-----GHLTVDTYGRPILEVPSVTPGPKGVNLCCTVDPLQGYTVLVK 55
DB 11 GCLLPILLLLFTGVSGEVSWFSVKGK---AEPITVLLGTEATLPCQLSPQSAARMHIR 66
QY 56 WLVRQSDPVTIFLRDSSGDHIQQAQYQGRHLVSHK---PGDVSLSLSTLEMDRSHVTC 113
DB 67 WYRAQPTPAVLVFNHNGQEQEVQMPETRGRTQMVYRQAIMDMSVALQIQVQASDDGLYHC 126
QY 114 EVTWQTPDG--NQVVRDKITELRV 135
DB 127 QFT----DGFTSQEVS---NELRV 143

RESULT 9
O8WMV3 PRELIMINARY; PRT; 365 AA.
ID O8WMV3
AC O8WMV3
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Coxsackie virus and adenovirus receptor BCAR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21547769; PubMed=11688979;
RA Thelen I., Keyaerts E., Lindberg M., Van Ranst M.;
RT "Characterization of a cDNA encoding the bovine coxsackie and
RT adenovirus receptor."
RL Biochem. Biophys. Res. Commun. 288:805-808(2001).
DR EMBL: AY033651; AAK57804.1;
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 2.
DR SMART: SM00409; Ig; 2.
DR SMART: SM00408; Igc2; 2.
DR SMART: SM00410; Ig_Like; 1.
KW Receptor.
SQ SEQUENCE 365 AA; 40153 MW; 36DE0BE5DCF8CF9 CRC64;

Query Match 11.2%; Score 104.5; DB 6; Length 365;
Best Local Similarity 27.6%; Pred. No. 0.055;
Matches 56; Conservative 25; Mismatches 81; Indels 41; Gaps 12;

QY 1 MGILLGLLLGLHGLTVDTYGRPILEVPSVTPGPKGVNLCCTVD---DPLQGYTVQ 52
DB 1 MELLRLFLLLCGVADFTRLSI--TTPQMIKAKAGAYLPCKFTLGPEDGGLD----- 54
QY 53 LVKWLVRGSGS---DPVTIFLRDSSGDHIQQAQYQ---GRHLVSHK---VPGDVSLSLSTL 103
DB 55 -IEWLLSPADNQKVDQVILY---SGDKIYDDYQDLKGRVHFHTSNDLKSQDASINTNL 110
QY 104 EMDRSHVTCVTTWQTPDGNQ---VVRDKITELRVKHSSKLL---KTKTEAPTMTY 155
DB 111 QLSDIGTQCKVKAPGVGNKKIQLTVLRPVGIRGICVVDGSEIGNDFKLKCE--PKGSL 169
```



QY 156 PL-----KATSTVRQSWDWTDM 173  
 ||| : : : : :  
 Db 170 PLRYEWQLSDSQKLPSTSWLPDM 192

## RESULT 10

Q9NQS3 Q9NQS3 PRELIMINARY; PRT; 549 AA.  
 AC Q9NQS3; 2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE Nectin 3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Raymond N., Borg J.-P., Lecoq E., Adelaide J., Campadelli-Fiume G.,  
 RA Dubreuil P., Lopez M.;  
 RT "Human nectin 3/PRR3: A novel member of the PVR/PRR/nectin family, that  
 RT interacts with afadin."  
 RL Gene 0:0-0(2000).  
 DR EMBL: AF282874; AAF97597.1; -  
 DR InterPro; IPR003599; Ig-like.  
 DR InterPro; IPR003600; Ig-like.  
 DR Pfam; PF00047; Ig; 2.  
 DR SMART; SM00409; Ig; 1.  
 DR SMART; SM00410; IG\_Like; 1.  
 SQ SEQUENCE 549 AA; 61002 MW; 6D1104CCB49D731 CRC64;

Query Match 10.9%; Score 102; DB 4; Length 549;  
 Best Local Similarity 26.0%; Pred. No. 0.17;  
 Matches 46; Conservative 22; Mismatches 95; Indels 14; Gaps 7;

QY 3 ILGLLLGLLTVTYGRPILEVPEVSTGPKGVDVSLQSLTLEMDRSHYTCVTVQVLPVQVRS 62  
 ||| : : : : :  
 Db 42 LLFLLFLLSRCLGALAG-PIIVPEH-VTVAVGKNSLKLIEVNETIQ--ISWEKIHGK 97  
 : : : : :  
 QY 63 DPVTFILRDSGDHIQQAQYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVTVQVLPD 121  
 : : : : :  
 Db 98 SSQVAVHHPQYGSVQYQYQGRVLFKNYSLNDATITLHNGFSDSGKYCKAVTF--PL 155  
 : : : : :  
 QY 122 GNQVYRDKITEL-----RVQKSSKLLTKTAPMTYPLKATSTVKQSDWTTDM 173  
 ||| : : : : :  
 Db 156 GNAQSSTTVTVLVEPTVSLIRGPDSLIDGNE--TVAACIAATGKPVAHIDWEGDL 210  
 : : : : :

## RESULT 11

Q9YGH1 Q9YGH1 PRELIMINARY; PRT; 335 AA.  
 AC Q9YGH1; 1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE Chrl thymocyte antigen precursor.  
 GN Chrl.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RPRL LINE 0; TISSUE=THYMUS;  
 RA Katevuo K.H., Boyd R., Gobel T.T., Bean A., Dunon D., Imhof B.A.,  
 RA Vainio O.;  
 RT "Chrl, a new Igsf member inhibits thymocyte differentiation at the  
 RT double positive stage."  
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; Y14063; CAA74390.1; -  
 DR HSP; P06907; INEU.  
 DR InterPro; IPR003600; Ig\_Like.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; Ig; 2.  
 DR SMART; SM00406; IG; 1.  
 DR SMART; SM00410; IG\_Like; 1.  
 KW Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 335 CHL1 THYMOCYTE ANTIGEN.  
 SQ SEQUENCE 335 AA; 36553 MW; AA640C5CD02CB16D CRC64;

Query Match 10.8%; Score 101; DB 13; Length 335;

Best Local Similarity 26.9%; Pred. No. 0.11;  
 Matches 36; Conservative 22; Mismatches 60; Indels 16; Gaps 7;

QY 9 LLGHLTVTYGRPILEVPEVSTGPKGVDVSLQSLTLEMDRSHYTCVTVQVLPVQVRS 62  
 ||| : : : : :  
 Db 15 LAGHV-----HG-VVTVPEKTVNVKGTG---GNATLLCTYTSSQPLGNFFIQWSFYSAKES 66  
 : : : : :  
 QY 63 DPVTFILRDSGDHIQQAQYQGRHLVSHKVPDGVSLQSLTLEMDRSHYTCVTVQVLPD 122  
 ||| : : : : :  
 Db 67 QLHTYIYY-SEGQSYSGEFGKDRITAATS-PGNASITISNNQPSDTGTCVCEVSPQDDA 124  
 : : : : :  
 QY 123 NQVYRDKITELRVQ 136  
 : : : : :  
 Db 125 QSQSKSVIVNVLVK 138  
 : : : : :

## RESULT 12

Q9PWR4 Q9PWR4 PRELIMINARY; PRT; 335 AA.  
 AC Q9PWR4; 2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE Chrl thymocyte antigen precursor.  
 GN Chrl.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H.B19; TISSUE=THYMUS;  
 RA Katevuo K.H., Boyd R., Gobel T.T., Bean A., Dunon D., Imhof B.A.,  
 RA Vainio O.;  
 RT "Chrl, a new Igsf member inhibits thymocyte differentiation at the  
 RT double positive stage."  
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

Query Match 10.8%; Score 101; DB 13; Length 335;

Best Local Similarity 26.9%; Pred. No. 0.11;  
 Matches 36; Conservative 22; Mismatches 60; Indels 16; Gaps 7;

QY 9 LLGHLTVTYGRPILEVPEVSTGPKGVDVSLQSLTLEMDRSHYTCVTVQVLPVQVRS 62  
 ||| : : : : :  
 Db 15 LAGHV-----HG-VVTVPEKTVNVKGTG---GNATLLCTYTSSQPLGNFFIQWSFYSAKES 66  
 : : : : :  
 QY 9 LLGHLTVTYGRPILEVPEVSTGPKGVDVSLQSLTLEMDRSHYTCVTVQVLPVQVRS 62  
 ||| : : : : :  
 Db 15 LAGHV-----HG-VVTVPEKTVNVKGTG---GNATLLCTYTSSQPLGNFFIQWSFYSAKES 66  
 : : : : :

```
QY 63 DPVTFLRDSGDHIOQAQYOGRLHVKVPGDVSLQSLSTLEMDRSHYTCVETWQPDG 122
DB 67 QLHTIYYIY-SEGOSYSYGFKDRITAATS-PGNASITISNQPDSGTGTCVFSPODDA 124
QY 123 NOVVRDKITELRVQ 136
DB 125 GOSOKSVIVNVLVK 138

RESULT 13
Q8VIM1
ID Q8VIM1 PRELIMINARY; PRT; 282 AA.
AC Q8VIM1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE TIM1.
GN TIM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DBA/2; TISSUE-SPLEEN;
RA Mcintire J.J., Umetsu S.E., Akbari O., Potter M., Barsh G.S.,
RA Freeman G.J., Umetsu D.T., Dekruyff R.H.;
RT "Tapr", a major T cell regulatory locus that controls the development
RT of airway hyperreactivity, cosegregates with variants in a novel gene
RT family";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF399830; AAL35775.1;
DR MGD; MGI:2159680; Tmd1.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00409; IG; 1.
SQ SEQUENCE 282 AA; 30968 MW; 7D30EE0698F0AC5F CRC64;

Query Match 10.7%; Score 100.5; DB 11; Length 282;
Best Local Similarity 27.1%; Pred. No. 0.097;
Matches 54; Conservative 19; Mismatches 81; Indels 45; Gaps 11;

QY 6 GLLLLGHLTVDTYGRPILEVPEVSGVPGKGVNLPCTYDPLQGYTVLVKLVQVGRSDP- 64
DB 10 GLILLPGAVDSY----VEYKGVVGH-----VTLPCTYSTYRGITTC--W--GRGQCS 57
QY 65 -----VTFLRDSGDHIOQAQYOGRLHVKVPGDVSLQSLSTLEMDRSHYTCVET--- 116
DB 58 SACONTLIWNGHRVTVQKSSRYNLKGHISE---GDVSLTIENSVESDGLYCCRVETPG 114
QY 117 WQTPDGNQVVRDKITELRVQKSHKLLKTKTEAPTMT---TYPL 157
DB 115 WFNDQKVTFSLOVKEI-PRPPRPRTTATGRPTTISTRTHVPTSTRVSTSTPT 173
QY 158 KA-TSTVKQSDWTDMDG 175
DB 174 STHTWTHKPDWNGTVTSSG 192

RESULT 14
Q8VIM2
ID Q8VIM2 PRELIMINARY; PRT; 305 AA.
AC Q8VIM2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE TIM1.
GN TIM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-SPLEEN;
RA Mcintire J.J., Umetsu S.E., Akbari O., Potter M., Barsh G.S.,
RA Freeman G.J., Umetsu D.T., Dekruyff R.H.;
RT "Tapr", a major T cell regulatory locus that controls the development
RT of airway hyperreactivity, cosegregates with variants in a novel gene
RT family";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF399829; AAL35774.1;
DR MGD; MGI:2159680; Tmd1.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00409; IG; 1.
SQ SEQUENCE 305 AA; 33391 MW; 8F4EA38627FE85FB CRC64;

Query Match 10.7%; Score 100; DB 11; Length 305;
Best Local Similarity 28.9%; Pred. No. 0.12;
Matches 48; Conservative 16; Mismatches 60; Indels 42; Gaps 10;

QY 6 GLLLLGHLTVDTYGRPILEVPEVSGVPGKGVNLPCTYDPLQGYTVLVKLVQVGRSDP- 64
DB 10 GLILLPGAVDSY----VEYKGVVGH-----VTLPCTYSTYRGITTC--W--GRGQCS 57
QY 65 -----VTFLRDSGDHIOQAQYOGRLHVKVPGDVSLQSLSTLEMDRSHYTCVET--- 116
DB 58 SACONTLIWNGHRVTVQKSSRYNLKGHISE---GDVSLTIENSVESDGLYCCRVETPG 114
QY 117 WQTPDGNQVVRDKITELRVQKSHKLLKTKTEAPTMT---TYPLKATST 162
DB 115 WFND-----QKVT-----FSLQVRKEIPTRP--PTRPTTT 142

RESULT 15
Q9YGV5
ID Q9YGV5 PRELIMINARY; PRT; 335 AA.
AC Q9YGV5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Cht1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99077161; PubMed-9862345;
RA Chretien I., Marcuz A., Courtet M., Katevuo K., Valnio O., Heath J.K.,
RA White S.J., Du Pasquier L.;
RT "CTX", a Xenopus thymocyte receptor, defines a molecular family
RT conserved throughout vertebrates";
RL Eur. J. Immunol. 28:4094-4104(1998).
DR EMBL: AF061023; AADI7523.1;
DR HSP; P06907; INEU.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00406; IG; 1.
DR SMART; SM00410; IG_like; 1.
SQ SEQUENCE 335 AA; 36561 MW; 071A313C3C56DCA0 CRC64;

Query Match 10.7%; Score 100; DB 13; Length 335;
Best Local Similarity 26.1%; Pred. No. 0.14;
Matches 35; Conservative 21; Mismatches 62; Indels 16; Gaps 6;

QY 9 LLGHLTVDTYGRPILEVPEV---TCPWKGVNLPCTYDPLQGYTVLVK---LVORGS 62
DB 15 LAGHVR-----GVVTVPEKTVNVKGTG---GNATLLCTYTSSOPLFFQWFSYSKES 66
```

Qy 63 DPTIFLRDSSGDHIQAKYQGRHLVSHKVPQDVSLQSLSTLEMDDRSHYTCEVTWQPPDG 122  
| | | | | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 67 QLHTIIYY-SEGQSYSGEFDKDRITAATS-PCNASITISNNQPSDTSYTCEVFSPODDA 124  
Qy 123 NQVVRDKITELRVQ 136  
| : : : : : | : : : : : | : : : : : | : : : : : |  
Db 125 GQSQKSVIYNVLVK 138

Search completed: March 4, 2003, 15:02:42  
Job time : 14.9893 secs

XX  
PI Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;  
PI Gorgone GA, Baughn MR, Patterson C;

XX WPI: 2000-256642/22.  
 DR N-PSDB; AAO8041.  
 XX  
 PT New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis  
 XX  
 PS Claim 1; Page 68; 75pp; English.  
 XX  
 CC AAA08035 to AAA08042 encode the human protein transport-associated  
 CC molecules (PTAMs) given in AAY82317 to AAY82324. The PTAMs have  
 CC cytosolic, antiarthritic, antiasthmatic, immunosuppressant,  
 CC antiarteriosclerotic, antiallergic, antidiabetic, antipsoriatic,  
 CC antirheumatic, osteopathic, dermatological, antianemic, antiparasitic,  
 CC hepatotropic, antitumor, antinflammatory and antiHIV activities, and  
 CC regulate protein transport. PTAM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PTAM. PTAM polynucleotides are useful for  
 CC diagnosing conditions associated with PTAM, comprising detecting PTAM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the  
 CC biological sample. Diseases prevented, treated or diagnosed include cell  
 CC proliferative disorders such as cancers, immune disorders, secretory  
 CC disorders and other conditions associated with abnormal vesicle  
 CC trafficking, such as allergies, asthma, urticaria and autoimmune  
 CC haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as  
 CC a targeting or delivery mechanism for bringing pharmaceutical agents  
 CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related  
 CC disorders. PTAM, its catalytic or immunogenic fragments are useful for  
 CC drug screening using libraries of compounds. PTAM polynucleotides are  
 CC useful for generating hybridisation probes useful in mapping the  
 CC naturally occurring genomic sequences.  
 XX  
 SQ Sequence 142 AA;  
 Query Match 100.0%; Score 749; DB 21; Length 142;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-76;  
 Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATSLDFKTYVDQACRAAEFVNIYETMDKRRALRLYLDKATLIWNGNAVSGLDALN 60  
 DB 1 MATSLDFKTYVDQACRAAEFVNIYETMDKRRALRLYLDKATLIWNGNAVSGLDALN 60  
 QY 61 NFFDTLPSEFQVNMDCQPVHEQATQSOTTVLVVTSGTVKFDGKQHFNFQFLTLTAQS 120  
 DB 61 NFFDTLPSEFQVNMDCQPVHEQATQSOTTVLVVTSGTVKFDGKQHFNFQFLTLTAQS 120  
 QY 121 TPNTVWKIASDCFRFQDWS 142  
 DB 121 TPNTVWKIASDCFRFQDWS 142

RESULT 2  
 ABB04459  
 ID ABB04459 standard; Protein: 142 AA.  
 AC ABB04459;  
 XX  
 DT 04-MAR-2002 (first entry)  
 XX  
 DE Human NTF2 associated protein 16.  
 XX  
 KW Human: NTF2 associated protein 16; cancer; HIV infection; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN CN1313332-A.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 15-MAR-2000; 2000CN-0114918.  
 XX  
 PR 15-MAR-2000; 2000CN-0114918.

XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
 XX PA  
 XX Mao Y, Xie Y;  
 PI  
 XX WPI: 2002-042199/06.  
 DR N-PSDB; ABA05754.  
 XX  
 PT Polypeptide-human NTF2 associated protein 16 and polynucleotide  
 PT encoding it  
 XX  
 PS Claim 1; Page 26(Disclosure); 32pp; Chinese.  
 XX  
 CC The present invention provides the protein and coding sequences of human  
 CC NTF2 associated protein 16. The sequences can be used in the treatment of  
 CC cancer and HIV infection, as well as other diseases. The present sequence  
 CC is the protein of the invention.  
 XX  
 SQ Sequence 142 AA;  
 Query Match 100.0%; Score 749; DB 23; Length 142;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-76;  
 Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATSLDFKTYVDQACRAAEFVNIYETMDKRRALRLYLDKATLIWNGNAVSGLDALN 60  
 DB 1 MATSLDFKTYVDQACRAAEFVNIYETMDKRRALRLYLDKATLIWNGNAVSGLDALN 60  
 QY 61 NFFDTLPSEFQVNMDCQPVHEQATQSOTTVLVVTSGTVKFDGKQHFNFQFLTLTAQS 120  
 DB 61 NFFDTLPSEFQVNMDCQPVHEQATQSOTTVLVVTSGTVKFDGKQHFNFQFLTLTAQS 120  
 QY 121 TPNTVWKIASDCFRFQDWS 142  
 DB 121 TPNTVWKIASDCFRFQDWS 142

RESULT 3  
 AAB95379  
 ID AAB95379 standard; Protein: 197 AA.  
 AC AAB95379;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:17707.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 XX  
 PR 27-AUG-1999; 99JP-0300253.  
 XX  
 PR 11-JAN-2000; 2000JP-0118776.  
 XX  
 PR 02-MAY-2000; 2000JP-0183767.  
 XX  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI: 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 17707; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 197 AA;

Query Match 100.0%; Score 749; DB 22; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-76;  
 Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATSLDFKTYVDQACRAAEFNYIYETMDKRRALTRYLDKATLIWNGNAVSLDALN 60  
 DB 56 MATSLDFKTYVDQACRAAEFNYIYETMDKRRALTRYLDKATLIWNGNAVSLDALN 115  
 QY 61 NFFDILPSEFVNMLDCOPVHEQATQSTTVLVTSYGVKFDGNKHQFFNFLTAAOS 120  
 DB 116 NFFDILPSEFVNMLDCOPVHEQATQSTTVLVTSYGVKFDGNKHQFFNFLTAAOS 175  
 QY 121 TPNNTVWKIASDCFRQDWSSS 142  
 DB 176 TPNNTVWKIASDCFRQDWSSS 197

RESULT 4  
 AAU16227  
 XX AAU16227 standard; Protein; 162 AA.  
 XX AC AAU16227;  
 XX DT 07-NOV-2001 (first entry)  
 XX DE Human novel secreted protein, Seq ID 1180.  
 XX DE Human; immunosuppressive; antiarthritic; antirheumatic;  
 KW cytotropic; cardiant; vasotropic; cerebroprotective; neotropic;  
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
 KW vulnary; secreted protein; rheumatoid arthritis;  
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
 KW corneal infection; wound healing; epithelial cell proliferation;  
 KW skin ageing; food additive; preservative; antiproliferative.  
 XX OS Homo sapiens.  
 XX PN WO200155322-A2.  
 XX PD 02-AUG-2001.  
 XX PD 17-JAN-2001; 2001WO-US01341.

XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0215647.  
 PR 11-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.

02-OCT-2000; 2000US-0236802.  
02-OCT-2000; 2000US-0237037.  
02-OCT-2000; 2000US-0237038.  
02-OCT-2000; 2000US-0237039.  
02-OCT-2000; 2000US-0237040.  
13-OCT-2000; 2000US-0239935.  
13-OCT-2000; 2000US-0239937.  
20-OCT-2000; 2000US-0240960.  
20-OCT-2000; 2000US-0241221.  
20-OCT-2000; 2000US-0241785.  
20-OCT-2000; 2000US-0241787.  
20-OCT-2000; 2000US-0241808.  
20-OCT-2000; 2000US-0241809.  
20-OCT-2000; 2000US-0241826.  
01-NOV-2000; 2000US-0244617.  
08-NOV-2000; 2000US-0246474.  
08-NOV-2000; 2000US-0246475.  
08-NOV-2000; 2000US-0246476.  
08-NOV-2000; 2000US-0246477.  
08-NOV-2000; 2000US-0246478.  
08-NOV-2000; 2000US-0246523.  
08-NOV-2000; 2000US-0246524.  
08-NOV-2000; 2000US-0246525.  
08-NOV-2000; 2000US-0246526.  
08-NOV-2000; 2000US-0246527.  
08-NOV-2000; 2000US-0246528.  
08-NOV-2000; 2000US-0246532.  
08-NOV-2000; 2000US-0246533.  
08-NOV-2000; 2000US-0246609.  
08-NOV-2000; 2000US-0246610.  
08-NOV-2000; 2000US-0246611.  
08-NOV-2000; 2000US-0246613.  
17-NOV-2000; 2000US-0249207.  
17-NOV-2000; 2000US-0249208.  
17-NOV-2000; 2000US-0249209.  
17-NOV-2000; 2000US-0249210.  
17-NOV-2000; 2000US-0249211.  
17-NOV-2000; 2000US-0249212.  
17-NOV-2000; 2000US-0249213.  
17-NOV-2000; 2000US-0249214.  
17-NOV-2000; 2000US-0249215.  
17-NOV-2000; 2000US-0249216.  
17-NOV-2000; 2000US-0249217.  
17-NOV-2000; 2000US-0249218.  
17-NOV-2000; 2000US-0249219.  
17-NOV-2000; 2000US-0249245.  
17-NOV-2000; 2000US-0249246.  
17-NOV-2000; 2000US-0249264.  
17-NOV-2000; 2000US-0249265.  
17-NOV-2000; 2000US-0249297.  
17-NOV-2000; 2000US-0249299.  
17-NOV-2000; 2000US-0249300.  
01-DEC-2000; 2000US-0250160.  
01-DEC-2000; 2000US-0250391.  
05-DEC-2000; 2000US-0251030.  
05-DEC-2000; 2000US-0251988.  
05-DEC-2000; 2000US-0256719.  
06-DEC-2000; 2000US-0251479.  
06-DEC-2000; 2000US-0251856.  
08-DEC-2000; 2000US-0251868.  
08-DEC-2000; 2000US-0251869.  
08-DEC-2000; 2000US-0251989.  
08-DEC-2000; 2000US-0251990.  
11-DEC-2000; 2000US-0254097.  
05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI: 2001-488783/53.  
N-PSDB; AAS26214.  
New nucleic acid molecules encoding 461 human secreted proteins for

diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives -  
Claim 11; SEQ ID No 1180; 980pp; English.  
The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present sequence represents a novel secreted protein of the invention.  
Query Match 97.1%; Score 727; DB 22; Length 162;  
Best Local Similarity 100.0%; Pred. No. 1.6e-73;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 DFTYVDQACRAAEFFNYIYETMDKRRRLTRLYDKATLIWNGNAVSGLDALNNFFDT 65  
DB 26 DFTYVDQACRAAEFFNYIYETMDKRRRLTRLYDKATLIWNGNAVSGLDALNNFFDT 85  
QY 66 LPSSEFQVNMDCQPVHEQATQSTTVLVVTSCTGVKDGNGKQHFNFQNLTAOSTPNT 125  
DB 86 LPSSEFQVNMDCQPVHEQATQSTTVLVVTSCTGVKDGNGKQHFNFQNLTAOSTPNT 145  
QY 126 VKTIASDCFRFQDWS 142  
DB 146 VKTIASDCFRFQDWS 162  
RESULT 5  
RAY82318  
ID RAY82318 standard; Protein: 140 AA.  
XX AAY82318;  
XX 19-JUN-2000 (first entry)  
DT Human protein transport molecule (PTAM) SEQ ID NO:2.  
XX Human; protein transport molecule; PTAM; diagnosis; cytostatic; antiarthritic; antiasthmatic; immunosuppressant; antiarteriosclerotic; antiallergic; antidiabetic; antilipemic; antirheumatic; osteopathic; dermatological; antianemic; antipsoriatic; hepatotropic; antitumor; antineoplastic; antihiv; protein transport regulator; cancer; immune disorder; cell proliferative disorder; secretory disorder; urticaria; allergy; abnormal vesicle trafficking; asthma; autoimmune haemolytic anaemia.  
XX Homo sapiens.  
OS WO200012703-A2.  
XX PN 09-MAR-2000.  
XX PD 26-AUG-1999; 99WO-US19616.  
XX PF

XX 27-AUG-1998; 98US-0098206.  
 XX (INCY-) INCYTE PHARM INC.  
 PA Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;  
 PI Gorgone GA, Baughn MR, Patterson C;  
 XX N-PSDB; AA08036.  
 DR WPI: 2000-256642/22.  
 XX New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis  
 XX Claim 1; Page 61-62; 75pp; English.  
 PS  
 CC AAA08035 to AAA08042 encode the human protein transport-associated  
 CC molecules (PTAMs) given in AA082317 to AA082324. The PTAMs have  
 CC cytosolic, antiarthritic, antilasthmatic, immunosuppressant,  
 CC antiarteriosclerotic, antiallergic, antidiabetic, antilipidemic,  
 CC antirheumatic; osteopathic dermatological, antianemic, antipsoriatic,  
 CC hepatotropic, antitumor, antinflammatory and antiHIV activities, and  
 CC regulate protein transport. PTAM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PTAM. PTAM polynucleotides are useful for  
 CC diagnosing conditions associated with PTAM, comprising detecting PTAM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the  
 CC biological sample. Diseases prevented, treated or diagnosed include cell  
 CC proliferative disorders such as cancers, immune disorders, secretory  
 CC disorders and other conditions associated with abnormal vesicle  
 CC trafficking, such as allergies, asthma, urticaria and autoimmune  
 CC haemolytic anemia. Anti-PTAM antibodies may be used as antagonists, as  
 CC a targeting or delivery mechanism for bringing pharmaceutical agents  
 CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related  
 CC disorders. PTAM, its catalytic or immunogenic fragments are useful for  
 CC drug screening using libraries of compounds. PTAM polynucleotides are  
 CC useful for generating hybridisation probes useful in mapping the  
 CC naturally occurring genomic sequences.  
 XX Sequence 140 AA;  
 SQ  
 Query Match 78.0%; Score 584; DB 21; Length 140;  
 Best Local Similarity 75.4%; Pred. No. 1.5e-57;  
 Matches 104; Conservative 22; Mismatches 12; Indels 0; Gaps 0;  
 QY 4 SLDFKTYVDAQRAAEFYNIYETMDKRRRLRLYLDKATILWNGNAVGLDALNFF 63  
 DB 3 SVDFKTYVDAQRAAEFYNIYETMDKRRRLRLYLDKATILWNGNAVGLDALNFF 62  
 QY 64 DTLPSSEFQVNLDCOPVHEQATQSTTVLVVTSVGVKEDGNKHQFFNQLTASTPN 123  
 DB 63 EMLPSEFQISVVDQCPVHEATPSQTTVLVWICSGVKEGNKQDFNQLTASTPN 122  
 QY 124 NTWKIASDCFRFDWSS 141  
 DB 123 NTWKIASDCFRFDWAS 140  
 RESULT 6  
 AAU16228.  
 ID AAU16228 standard; Protein; 164 AA.  
 XX  
 AC AAU16228;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX Human novel secreted protein, Seq ID 1181.  
 DE  
 XX Human; immunosuppressive; antiarthritic; antirheumatic;  
 KW cytostatic; cardiatic; vasotropic; cerebroprotective; neurotropic;  
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
 KW vulnary; secreted protein; rheumatoid arthritis;

KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
 KW corneal infection; wound healing; epithelial cell proliferation;  
 KW skin ageing; food additive; preservative; antiproliferative.  
 OS Homo sapiens.  
 XX WO200155322-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01341.  
 XX  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226682.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.





OS Drosophila melanogaster.  
 XX PN W0200171042-A2.  
 XX PD 27-SEP-2001.  
 XX XX 23-MAR-2001; 2001WO-US09231.  
 XX PF 23-MAR-2000; 2000US-191637P.  
 XX PR 11-JUL-2000; 2000US-0614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI; 2001-656860/75.  
 XX DR N-PSDB; ABL09120.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX PT interactions.  
 XX PS Disclosure; SEQ ID NO 21843; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX CC useful in developmental biology and in elucidating cell signalling and  
 XX CC cell-cell interactions in higher eukaryotes for the development of  
 XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 XX CC sequences (ABL01840-ABL16175) and the encoded proteins  
 XX CC (ABB57737-ABB72072).  
 XX CC The sequence data for this patent did not form part of the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 133 AA;  
 Query Match: 41.4%; Score 310; DB 22; Length 133;  
 Best Local Similarity: 42.9%; Pred. No. 7.9e-27;  
 Matches 57; Conservative 24; Mismatches 48; Indels 4; Gaps 1;  
 QY 6 DFTYVDOACRAAEFVNIYETMDKRRALTRYLDKATLIWNGNAVSGLDALNNEFD 65  
 Db 4 DLKAKVESCARTDFTRLYASVDNRQQIGRLYLDNATLSWNGAIGQMIESYFQE 63  
 QY 66 LPSEFQWMLDCOPVHEQATQSTTVLVVTSGTVKFGNKHQFNFQNLTAQSTENNT 125  
 Db 64 LPSSNHQNLWDQPIVDQAVSNQALYLINASGSKVADQQLRKFQTFIVTAE----ND 119  
 QY 126 VKTIASDCFRFQD 138  
 Db 120 KKVVSDCYRMQE 132  
 RESULT 8  
 AAG12073  
 ID AAG12073 standard; Protein; 125 AA.  
 XX AC AAG12073;  
 XX DT 17-OCT-2000 (first entry)  
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 11042.  
 XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 XX KW termination sequence.  
 XX OS Arabidopsis thaliana.  
 XX PN EP1033405-A2.  
 XX PD 06-SEP-2000.  
 XX PF 25-FEB-2000; 2000EP-0301439.  
 XX XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 25-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0132863.  
 PR 14-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138340.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.



XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0123788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149920.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.



XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL10155.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Disclosure; SEQ ID NO 24948; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 690 AA;  
 SQ  
 Query Match 15.18; Score 113; DB 22; Length 690;  
 Best Local Similarity 31.56; Pred. No. 0.00089;  
 Matches 39; Conservative 14; Mismatches 55; Indels 16; Gaps 5;  
 QY 20 EFVNYYETMDKRRRLRLYLDKATLI-WNGNAVYGLDALNNFFDTLPSSSEFOVNMDC 78  
 DB 19 EFVRYVYLLNKAPNHLRFFYNHNSYTHGSKLVVGOEIHNRQ-----QLNFND 71  
 QY 79 ----QPVEHQATQSTTVLVVTSVGVKFDGNKQHFNFQNLTAOSTPNNVTWKIASDCF 134  
 DB 72 HAKISQVDAQATLGNVGVVQVT-GELSDNGQPMRRTQTQTVLAQSPKK---YYVHNDIF 127  
 QY 135 RFQD 138  
 DB 128 RYQD 131  
 RESULT 12  
 AAG08682  
 ID AAG08682 standard; Protein; 123 AA.  
 XX AC  
 XX AAG08682;  
 XX DT  
 XX 17-OCT-2000 (first entry)  
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 6317.  
 XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX OS Arabidopsis thaliana.  
 XX PN EPI033405-A2.  
 XX PD 06-SEP-2000  
 XX PF 25-FEB-2000; 2000EP-0301439.  
 XX PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126788.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 08-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 23-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159394.

PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160988.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 112.5; DB 21; Length 123;

Best Local Similarity 27.3%; Pred. No. 0.0001;

Matches 33; Conservative 25; Mismatches 52; Indels 11; Gaps 5;

QY 18 AEEFVNIYYETMDKRRALTRYLDKATLIWNGNAVSGLDALNNFFDTLP--SSEFQVNM 75

Db 7 SRAFVEHYISTFTDNRVGLAGLYQEAASMTLTFEGKIQGVQSIKVLTSLPFOCKHHIST 66

QY 76 LDCQPVHEQATQSTTVLVYVTSKTVKFDGNKQHF-FNQNFLLTAQSTPNNTVWKIASDCF 134

Db 67 VDCQP-----SGPASGMLVFVSGNLQAGEHALKFSQMFHL--MPTPGGSFY-VFNDIF 118

QY 135 R 135

Db 119 R 119

RESULT 13

AAG08681

ID AAG08681 standard; Protein; 126 AA.

XX

AC AAG08681;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 6316.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0156559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.



```

PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 112.5; DB 21; Length 126;
Best Local Similarity 27.3%; Pred. No. 0.00011;
Matches 33; Conservative 25; Mismatches 52; Indels 11; Gaps 5;

Qy 18 AEEFNIIYETMDRRRRALRLYLDKATLIWGNVAVSCLDALNNFFDTLP--SSEFQVNM 75
Db 10 SKAFVHYIYFTDNRVGLAGLYQESASMLTFEGQKIQGVQIVAKLSLPQQCKHHIST 69
Qy 76 LDCOPVHEQATQSTTVLVVTSQTVKFDGNQHF-FNQNLITAQSTENNVTWVKIASDCF 134
Db 70 VDCQP-----SGPASGMLVFFSGNLQLAGEEHAKFSQMFHL--MPTQGSFY-VFNDIF 121

Qy 135 R 135
Db 122 R 122

RESULT 14
AAG05147
ID AAG05147 standard; Protein; 131 AA.
XX
AC AAG05147;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1443.
XX
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.

```

```

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.

```

```
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 112.5; DB 21; Length 131;
Best Local Similarity 26.8%; Pred. No. 0.00011;
Matches 33; Conservative 26; Mismatches 53; Indels 11; Gaps 5;

QY 18 AEEFNIYYETMDKRRALTRYLDKATLIWNGNAVSGLDALNNFFDTLP--SSEFOVNM 75
DB 7 SKAFVEHYSTFTNRYGLAGLYQEASMLTFEGQKIQGVQSIKVLTSLPFOQCKHHIST 66
QY 76 LDCQPVHEQATQSTTVLVVTSGTVKFDGKQHF-FNQNFLTAAQSTPNTNTYKIASDCF 134
DB 67 VDCQP-----SGPASGMLVTVSGNLQAGEKHALKFSQMFHL--MPTPGSFY-VFNDIF 118

QY 135 RFQ 137
DB 119 SWR 121

RESULT 15
AAG05146
ID AAG05146 standard; Protein; 134 AA.
XX
AC AAG05146;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1442.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
```

PR 29-MAR-1999; 99US-0136785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142377.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

```
PR 12-OCT-1999; 99US-0158369;
PR 13-OCT-1999; 99US-0158293;
PR 13-OCT-1999; 99US-0159294;
PR 13-OCT-1999; 99US-0159295;
PR 14-OCT-1999; 99US-0159329;
PR 14-OCT-1999; 99US-0159330;
PR 14-OCT-1999; 99US-0159331;
PR 14-OCT-1999; 99US-0159637;
PR 14-OCT-1999; 99US-0159638;
PR 18-OCT-1999; 99US-0159584;
PR 21-OCT-1999; 99US-0160741;
PR 21-OCT-1999; 99US-0160767;
PR 21-OCT-1999; 99US-0160768;
PR 21-OCT-1999; 99US-0160770;
PR 21-OCT-1999; 99US-0160814;
PR 21-OCT-1999; 99US-0160815;
PR 22-OCT-1999; 99US-0160980;
PR 22-OCT-1999; 99US-0160981;
PR 22-OCT-1999; 99US-0160989;
PR 25-OCT-1999; 99US-0161404;
PR 25-OCT-1999; 99US-0161405;
PR 25-OCT-1999; 99US-0161406;
PR 26-OCT-1999; 99US-0161359;
PR 26-OCT-1999; 99US-0161360;
PR 26-OCT-1999; 99US-0161361;
PR 28-OCT-1999; 99US-0161920;
PR 28-OCT-1999; 99US-0161992;
PR 28-OCT-1999; 99US-0161993;
PR 29-OCT-1999; 99US-0162142;

Query Match : 15.0%; Score 112.5; DB 21; Length 134;
Best Local Similarity 26.8%; Pred. No. 0.00012;
Matches 33; Conservative 26; Mismatches 53; Indels 11; Gaps 5;

QY 18 AEEFNYYETDKRRALTRYLDKATLIWNGNAVSGLDALNFFDTLP--SSEFQVNM 75
Db :
10 SKAFVEHYTSTFTNRVGLAGLYQEASMLTFEGKIQGVQSIYAKLTSLPFOCKHHIST 69
QY 76 LDCQPVEQATQSOTVLVVTSGTVKFDGNKQHF-FNQNFLLTAQSTPNNVTYKIASDCF 134
Db :
70 VDCQP-----SGPASGMLYFVSGNLQAGEKHALKESQMFHL--MPTPGSPY-VPNDIF 121
QY 135 REQ 137
Db : :
122 SWR 124
```

Search completed: March 4, 2003, 14:58:48  
Job time : 13.2741 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:55:07 ; Search time 4.92282 seconds  
(without alignments)  
848.711 Million cell updates/sec

Title: US-09-763-902B-7

Perfect score: 749

Sequence: 1 MATSLDKTYVDQACRAAE.....NNTVWKIASDCFRFDWSSS 142

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| 1          | 76    | 10.1        | 238    | 4     | US-08-858-207A-420  |
| 2          | 71.5  | 9.5         | 748    | 4     | US-08-392-459-24    |
| 3          | 71.5  | 9.5         | 748    | 4     | US-08-392-459-28    |
| 4          | 71.5  | 9.5         | 748    | 4     | US-08-392-459-34    |
| 5          | 71.5  | 9.5         | 748    | 5     | PCT-US91-08525-24   |
| 6          | 71.5  | 9.5         | 748    | 5     | PCT-US91-08525-28   |
| 7          | 71.5  | 9.5         | 748    | 5     | PCT-US91-08525-34   |
| 8          | 71.5  | 9.5         | 748    | 5     | PCT-US91-08525-38   |
| 9          | 71.5  | 9.5         | 748    | 5     | PCT-US93-04384-6    |
| 10         | 71.5  | 9.5         | 1454   | 4     | US-08-392-459-22    |
| 11         | 71.5  | 9.5         | 1454   | 4     | US-08-392-459-26    |
| 12         | 71.5  | 9.5         | 1454   | 5     | PCT-US91-08525-22   |
| 13         | 71.5  | 9.5         | 1454   | 5     | PCT-US91-08525-28   |
| 14         | 71.5  | 9.5         | 1454   | 5     | PCT-US93-04384-2    |
| 15         | 71.5  | 9.5         | 1454   | 5     | PCT-US93-04384-8    |
| 16         | 71.5  | 9.5         | 1454   | 5     | PCT-US93-04384-16   |
| 17         | 71.5  | 9.5         | 1454   | 5     | PCT-US93-04384-44   |
| 18         | 71.5  | 9.5         | 1454   | 5     | PCT-US93-04384-45   |
| 19         | 71.5  | 9.5         | 1454   | 5     | PCT-US93-04384-47   |
| 20         | 71.5  | 9.5         | 1939   | 4     | US-09-310-187A-1    |
| 21         | 70.5  | 9.4         | 474    | 4     | US-09-134-001C-3241 |
| 22         | 68.5  | 9.1         | 1335   | 1     | US-09-134-001C-3716 |
| 23         | 67.5  | 9.0         | 259    | 1     | US-08-403-866-5     |
| 24         | 67.5  | 9.0         | 1120   | 4     | US-09-147-404-1     |
| 25         | 67    | 8.9         | 544    | 3     | US-08-591-685-9     |
| 26         | 66.5  | 8.9         | 498    | 4     | US-09-238-303-8     |
| 27         | 66.5  | 8.9         | 1886   | 4     | US-08-938-103-3     |

Sequence 12, Appl  
Sequence 190, Appl  
Sequence 190, Appl  
Sequence 190, Appl  
Sequence 190, Appl  
Sequence 3, Appl  
Sequence 9, Appl  
Sequence 32, Appl  
Sequence 32, Appl  
Sequence 12, Appl  
Sequence 43, Appl  
Sequence 46, Appl  
Sequence 48, Appl  
Sequence 112, Appl  
Sequence 112, Appl  
Sequence 112, Appl  
Sequence 6, Appl  
Sequence 104, Appl

#### ALIGNMENTS

RESULT 1  
US-08-858-207A-420  
; Sequence 420, Application US/08858207A  
; Patent No. 6348328  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael  
; APPLICANT: Hodgson, John  
; APPLICANT: Knowles, David  
; APPLICANT: Nicholas, Richard  
; APPLICANT: Stodola, Robert  
; TITLE OF INVENTION: No. 6348328el Compounds  
; NUMBER OF SEQUENCES: 552  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/858,207A  
; FILING DATE: 09-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/017670  
; FILING DATE: 14-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmel, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P50475  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 420:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 238 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 6348328e  
US-08-858-207A-420

Query Match 10.1%; Score 76; DB 4; Length 238;  
Best Local Similarity 23.7%; Pred. No. 0.35;



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein-search, using sw model

Run on: March 4, 2003, 14:53:52 ; Search time 6.38795 Seconds  
(without alignments)  
2137.006 Million cell updates/sec

Title: US-09-763-902b-7  
Perfect score: 749  
Sequence: 1 MATSLDFKTYVDQACRAAE.....NNTVWKIASDCFRFQDWSSS 142

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:  
1: piri:  
2: pir2:  
3: pir3:  
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 115   | 15.4        | 127    | S00751    | placental protein  |
| 2          | 113   | 15.1        | 123    | T38039    | probable nuclear   |
| 3          | 112.5 | 15.0        | 126    | B86405    | probable nuclear   |
| 4          | 107.5 | 14.4        | 522    | T23J18.22 | protein T23J18.22  |
| 5          | 105   | 14.0        | 122    | H86248    | protein F17L21.10  |
| 6          | 101.5 | 13.6        | 125    | S50467    | nuclear transport  |
| 7          | 99.5  | 13.3        | 133    | T23921    | hypothetical prote |
| 8          | 99.4  | 12.6        | 537    | T23479    | hypothetical prote |
| 9          | 90.5  | 12.1        | 428    | F86270    | hypothetical prote |
| 10         | 86.5  | 11.5        | 643    | S24382    | nitrous-oxide redu |
| 11         | 81    | 10.8        | 427    | E96716    | probable RNA-bind  |
| 12         | 81    | 10.8        | 515    | S63382    | hypothetical prote |
| 13         | 81    | 10.8        | 872    | C72042    | conserved hypothet |
| 14         | 81    | 10.8        | 872    | D86581    | CT619 hypothetical |
| 15         | 80.5  | 10.7        | 423    | H84450    | hypothetical prote |
| 16         | 79.5  | 10.6        | 4688   | F82885    | hypothetical prote |
| 17         | 78.5  | 10.5        | 334    | T47693    | hypothetical prote |
| 18         | 78.5  | 10.5        | 1144   | A36968    | pi-like adhesin pr |
| 19         | 78    | 10.4        | 394    | T32042    | hypothetical prote |
| 20         | 77.5  | 10.3        | 910    | S40930    | hypothetical prote |
| 21         | 77    | 10.3        | 340    | G64029    | hypothetical prote |
| 22         | 77    | 10.3        | 579    | W2BE34    | gene 34 protein -  |
| 23         | 76.5  | 10.2        | 434    | T40805    | probable RNA-bind  |
| 24         | 76    | 10.1        | 418    | T27506    | hypothetical prote |
| 25         | 74.5  | 9.9         | 287    | E82881    | hypothetical prote |
| 26         | 74    | 9.9         | 287    | F57925    | DNA-formamidopyrim |
| 27         | 72.5  | 9.7         | 403    | H64861    | hypothetical prote |
| 28         | 72.5  | 9.7         | 528    | A97270    | Arpase component o |
| 29         | 72.5  | 9.7         | 557    | G86357    | Similar to auxin-1 |

30 72.5 9.7 1336 2 T18288  
31 72 9.6 563 2 G02043  
32 71.5 9.5 186 2 S77560  
33 71.5 9.5 374 2 T07033  
34 71.5 9.5 1452 1 VGIH79  
35 71.5 9.5 1939 1 A46762  
36 71 9.5 278 2 D82275  
37 71 9.5 680 2 G86672  
38 70.5 9.4 326 2 B84971  
39 70.5 9.4 567 2 T25771  
40 70.5 9.4 1039 2 G83748  
41 70 9.3 320 2 G86384  
42 70 9.3 631 2 S36505  
43 70 9.3 711 2 D91238  
44 70 9.3 711 2 H86085  
45 70 9.3 711 2 F65201

ALIGNMENTS

RESULT 1

S00751  
placental protein 15 - human  
N;Alternate names: Pp15  
C;Species: Homo sapiens (man)  
C;Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 24-Sep-1999  
C;Accession: S00751  
R;Grundmann, U.; Nerlich, C.; Rein, T.; Lottspeich, F.; Kuepper, H.A.  
Nucleic Acids Res. 16, 4721, 1988  
A;Title: Isolation of cDNA coding for the placental protein 15 (PP15).  
A;Reference number: S00751; MUID:88247772; PMID:3380696  
A;Accession: S00751  
A;Molecule type: mRNA  
A;Residues: 1-127 <GRU>  
A;Cross-references: EMBL:X07315; NID:g35578; PIDN:CAA30278.1; PID:g35579  
C;Superfamily: yeast nuclear transport factor NTF2

Query Match 15.4% Score 115; DB 2; Length 127;  
Best Local Similarity 27.4%; Pred. No. 0.00039;  
Matches 32; Conservative 18; Mismatches 55; Indels 12; Gaps 3;

Qy 21 FVNIYETMDKRRRALTRYLDKATLIWNGAVSGLDNLNNFFDTLPSEFQ--VNMLDC 78  
Db 14 FIQHYQLFNDRTOLGAIYIDASCLTWEGQFQGAIVKLSLPFKIQHSITAQDH 73  
Qy 79 QPVHEQATQSQTVLVWTSQTVKFGNKHQFNQFLITAQSTPNTVWKIASDCFR 135  
Db 74 QPTPD-----SCIISMVYVQLKADEDFIMGFHMFLL-----KNINDAWVCTNDMFR 120

RESULT 2

T38039  
probable nuclear transport factor 2 - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 15-Sep-2000  
C;Accession: T38039; T37728  
R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Whitehead, S.; Churcher, C.M.  
submitted to the EMBL Data Library, August 1999  
A;Reference number: Z21764  
A;Accession: T38039  
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA  
A;Residues: 1-123 <MCD>  
A;Cross-references: EMBL:AL109951; PIDN:CA853052.1; GSPDB:GN00066; SPDB:SPAC1B9.01c  
A;Experimental source: strain 972h; cosmid c1B9  
R;Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, December 1995  
A;Reference number: Z21740  
A;Accession: T37728  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: SL/77-123 <DEV>

A:Cross-references: EMBL:Z68197; PIDN:CRA92380.1; GSPDB:GN00066; SPDB:SPAC15F9.03c  
C:Genetics:  
A:Gene: SPDB:SPAC1B9.01c  
A:Map position: 1  
A:Introns: 3/1; 30/3; 54/3; 88/3  
C:Superfamily: yeast nuclear transport factor NTF2

Query Match 15.1%; Score 113; DB 2; Length 123;  
Best Local Similarity 28.1%; Pred. No. 0.00059;  
Matches 34; Conservative 24; Mismatches 51; Indels 12; Gaps 4;  
QY 18 AEEFVNIYETMDKRRRLRLYLKATLIWNGNAVSGLDALNNFFDTLPSSFEQ--VNM 75  
DB 8 ATQFTQFYQTFDSQSSQLSRYRESMSLTFEGAQLOGTKAIVKLVSLPFRVQHRIST 67  
QY 76 LDCQPVHEQATQSTTVLVVTSYGVKFDGKQHF--FNQNFLLTAQSTPNTVTKWIASDC 134  
DB 68 LDAQP-----TGTGTVIMVTGELLDEEQNAQRYSQVFL-----VNNNGYVYVNDLF 118  
QY 135 R 135  
DB 119 R 119

RESULT 3  
B86405  
probable nuclear transport factor 2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: B86405  
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A:Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: B86405  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-126 <STO>  
A:Cross-references: GB:AE005172; NID:g11024877; PIDN:AG26961.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
C:Superfamily: yeast nuclear transport factor NTF2

Query Match 15.0%; Score 112.5; DB 2; Length 126;  
Best Local Similarity 27.3%; Pred. No. 0.00068;  
Matches 33; Conservative 25; Mismatches 52; Indels 11; Gaps 5;  
QY 18 AEEFVNIYETMDKRRRLRLYLKATLIWNGNAVSGLDALNNFFDTLP--SSEFQVNM 75  
DB 10 SKAFVEHYSTFDNRVGLAGLYEASMLTFEGQIQGVQSVAKLTSLPQCKKHIST 69  
QY 76 LDCQPVHEQATQSTTVLVVTSYGVKFDGKQHF--FNQNFLLTAQSTPNTVTKWIASDC 134  
DB 70 VDCQP-----SGPASGMLVFSGNLQAGEHALKFSOMFHL--MPTFGSFF--VFNDIF 121  
QY 135 R 135  
DB 122 R 122

RESULT 4  
H86248  
protein T23J18.22 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H86248

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H86248  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-522 <STO>  
A:Cross-references: GB:AE005172; NID:96554189; PIDN:AAF16635.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: T23J18.22  
A:Map position: 1

Query Match 14.4%; Score 107.5; DB 2; Length 522;  
Best Local Similarity 29.4%; Pred. No. 0.012;  
Matches 30; Conservative 17; Mismatches 50; Indels 5; Gaps 3;  
QY 18 AEEFVNIYETMDKRRRLRLYLKATLIWNGNAVSGLDALNNFFDTLPSSFEQ--VNM 75  
DB 340 ASAPVNHVHLFDNRSLSSLYNPTSLTFEGQIYGVNISKRLQLPDQCHHIST 399  
QY 76 LDCQPVHEQATQSTTVLVVTSYGVKFDGKQHF--FNQNFLL 116  
DB 400 VDSQP--SSMAGCGGILVFSGSIQLHGDPHLPFSQTFHL 439

RESULT 5  
H86398  
protein F17L21.10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H86398  
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H86398  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-122 <STO>  
A:Cross-references: GB:AE005172; NID:g9802547; PIDN:AAF99749.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F17L21.10  
A:Map position: 1  
C:Superfamily: yeast nuclear transport factor NTF2

Query Match 14.0%; Score 105; DB 2; Length 122;  
Best Local Similarity 29.5%; Pred. No. 0.0035;  
Matches 36; Conservative 20; Mismatches 52; Indels 14; Gaps 5;  
QY 18 AEEFVNIYETMDKRRRLRLYLKATLIWNGNAVSGLDALNNFFDTLPSSFEQVNM-- 75  
DB 7 AKAFVEHYSTFDANRPLGLVSLYQESMLTFEGQIQGVQSVAKLTLGLPQCKKHIST 66  
QY 76 LDCQPVHEQATQSTTVLVVTSYGVKFDGKQHF--FNQNFLLTAQSTPNTVTKWIASDC 133  
DB 67 VDCQP-----SGPAGMLVFSGNLQLAG--EOHALKFSQFHLIS-----NOGNYVFNDI 116  
QY 134 FR 135



Db 117 FR 118

RESULT 6

S50467

nuclear transport factor NTF2 - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein YER009w

C:Species: *Saccharomyces cerevisiae*

C:Date: 28-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 24-Sep-1999

C:Accession: S50467; S41793; S72327

R:Dietrich, F.S.

submitted to the EMBL Data Library, December, 1994

A:Description: The sequence of *S. cerevisiae* cosmid 9537, 9581, 9495, 9867, and lambda

A:Reference number: S50459

A:Accession: S50467

A:Molecule type: DNA

A:Residues: 1-125 <DIE>

A:Cross-references: EMBL:U18778; NID:g603592; PIDN:AAB64542.1; PID:g603601; MIPS:YER009w

R:Haarer, B.K.; Petzold, A.S.; Brown, S.S.

submitted to the EMBL Data Library, July 1993

A:Description: Identification of mutations that are synthetically lethal with altered ye

A:Reference number: S41793

A:Accession: S41793

A:Molecule type: DNA

A:Residues: 1-84 <HAA>

A:Cross-references: EMBL:L22204; NID:g347714; PIDN:AAB49379.1; PID:g347715

R:Haarer, B.K.; Corbett, A.; Kweon, Y.; Petzold, A.S.; Silver, P.; Brown, S.S.

Genetics 144, 495-510, 1996

A:Title: SEC3 mutations are synthetically lethal with profilin mutations and cause defec

A:Reference number: S72327; MUID:97044444; PMID:889515

A:Accession: S72327

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-84 <HAA>

A:Cross-references: EMBL:L22204; NID:g347714; PIDN:AAB49379.1; PID:g347715

C:Genetics:

A:Gene: SGD:NTF2

A:Cross-references: SGD:S0000811; MIPS:YER009w

A:Map position: 5R

C:Superfamily: yeast nuclear transport factor NTF2

Query Match 13.6%; Score 101.5; DB 2; Length 125;

Best Local Similarity 27.9%; Pred. No. 0.008; Mismatches 58; Indels 21; Gaps 6;

Matches 38; Conservative 19;

QY 4 SLDFYTVDOACRAAEFVNIYETMDKRRALRLYLKATLIWNGNAVSGLDALNNFF 63

Db 2 SLDFNT-----LAQNFQFYNGFDTRDSQGLNLYRNESMLTETSQLOKAKDIVEKL 54

QY 64 DTLPSSEFQ--VNMLDCQPVHEQATQSTVLVVTSGTVKFD--GNKQHFNFNLLTAQ 119

Db 55 VSLPFQVKVQHRITTLDAQAPSPG-----DVLVMTGDLLEDEQNQR-FSQVFLH--- 105

QY 120 STPNNTVWKIADCFR 135

Db 106 -IPDGNSTYVFNDIFR 120

RESULT 7

T23921

hypothetical protein R05D11.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

R:White, S.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19818

A:Accession: T23921

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-133 <WIL>

A:Cross-references: EMBL:Z75546; PIDN:CAA99890.1; GSPDB:GN00019; CESP:R05D11.3

A:Experimental source: clone R05D11

C:Genetics:

A:Gene: CESP:R05D11.3

A:Map position: 1

A:Introns: 33/1; 63/3; 96/3

C:Superfamily: yeast nuclear transport factor NTF2

Query Match 13.3%; Score 99.5; DB 2; Length 133;

Best Local Similarity 26.8%; Pred. No. 0.014;

Matches 34; Conservative 27; Mismatches 47; Indels 19; Gaps 6;

QY 18 AEEFVNIYETMD-----KRRRALRLYL-LDKATLIWNGNAVSGLDALNNFFDLPSEF 71

Db 11 AKAFIQHYYSKFDVGDGMSRAQGLSDLYDPENSYMTTEGQQAQKGRDGLQKFTTLGFTKI 70

QY 72 Q--VNMLDCQPVHEQATQSTVLVVTSGTVKFDGNKQHFNFNLLTAQSTPNNT-VWK 128

Db 71 QRAITVIDSQPLDGSIQ-----VMVLGQLKTDDEDPINPFSQVFL----RPNNOGSYF 120

QY 129 IASDCFR 135

Db 121 IGNEIFR 127

RESULT 8

T23479

hypothetical protein K08F4.2 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T23479

R:Hembry, C.

submitted to the EMBL Data Library, January 1996

A:Reference number: Z19746

A:Accession: T23479

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-537 <WIL>

A:Cross-references: EMBL:Z68879; PIDN:CAA93082.1; GSPDB:GN00022; CESP:K08F4.2

A:Experimental source: clone K08F4

C:Genetics:

A:Gene: CESP:K08F4.2

A:Map position: 4

A:Introns: 66/3; 99/1; 140/3; 161/3; 338/1; 419/1

Query Match 12.6%; Score 94; DB 2; Length 537;

Best Local Similarity 25.0%; Pred. No. 0.26;

Matches 32; Conservative 28; Mismatches 48; Indels 20; Gaps 6;

QY 17 AAEE-----FVNIYETMDKRRALRLYLKATLIWNGNAVSGLDALNNFFDTLP-SSSEF 71

Db 96 AAEOVGGAFCQFYITVSENRAIKTFYGHESKFYLDQDVVTGSOEIVKLYNHLPETTHF 155

QY 72 QVNMLDCQPV-HEQATQSTVLVVTSGTVKFDGNKQHFNFNLLTAQSTPNNTVWKIA 130

Db 156 KIOSIKGYTPHKQG-----VIINVIGTVNL-----RPFLOSFLLGQGGQKK---YYVE 201

QY 131 SDCFRQD 138

Db 202 SDAFOYID 209

RESULT 9

F86270

hypothetical protein F21F23.16 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C:Accession: F86270

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: F86270  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-428 <STO>  
 A:Cross-references: GB:AE005172; NID:98920577; PIDN:AAF81299.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 12.1%; Score 90.5; DB 2; Length 428;  
 Best Local Similarity: 24.8%; Pred. No. 0.43;  
 Matches 30; Conservative 25; Mismatches 53; Indels 13; Gaps 4;  
 QY 21 FVNIYETMDKRRALTRLYLDKATLIWNGN-----AVSGLDALNNFFDLPSPSEFQVNM 75  
 Db 19 FVEKYNNLLKSPSQVHQFVLDSDVGLRPGSGDGMVSVKSLKAIN---EQIMSFDFEISK 75  
 QY 76 LDCOPVHEQATQSTTVLVVTSQVTFKDGKQHFENQNFLLTAQSTPNNTVWKIADCFR 135  
 Db 76 IQILTADSQASYNGVYVTLVGLTIVKEGORMR-FSQSFFL-----VPLNGSYFVLNDVFR 130  
 QY 136 F 136  
 Db 131 Y 131

RESULT 10  
 S24382  
 nitrous-oxide reductase (BC 1.7.99.6) - Alcaligenes eutrophus  
 C:Species: Alcaligenes eutrophus  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S24382  
 R:Zunft, W.G.; Dreusch, A.; Loecheit, S.; Cuypers, H.; Friedrich, B.; Schneider, B.  
 Eur. J. Biochem. 208, 31-40, 1992  
 A:Title: Derived amino acid sequences of the nosZ gene (respiratory N(2)O reductase) from *Alcaligenes eutrophus*.  
 A:Reference number: S24382; MUID:92380183; PMID:1324835  
 A:Accession: S24382  
 A:Molecule type: DNA  
 A:Residues: 1-843 <ZUM>  
 A:Cross-references: EMBL:X65278; NID:938781; PIDN:CAA46383.1; PID:g38782  
 C:Superfamily: nitrous-oxide reductase  
 C:Keywords: copper; metalloprotein; oxidoreductase  
 F:589,624,628/Binding site: copper 1 (His, Cys, Cys) #status predicted  
 F:624,626,628,632/Binding site: copper 2 (Cys, His, Cys, His) #status predicted

Query Match 11.5%; Score 86.5; DB 1; Length 643;  
 Best Local Similarity: 31.7%; Pred. No. 1.7;  
 Matches 32; Conservative 10; Mismatches 46; Indels 13; Gaps 4;  
 QY 26 YETMDKRRALTRLYLDKATLIWNGVSGLDALNNFFDLPSPSEFQVNMDC--QPVEH 83  
 Db 387 HTAFDGRGNVYTLFLDSQVKN-----LDAIKFKHGDKNKKYVVDRLDLQVQPGHV 450  
 QY 84 QATQSTTVLVVTSQVTFKDGKQHFENQNFLLTAQSTPNN 124  
 Db 451 NASQSET---VAADGKYLAVGCK--FSKDRPLVGPLHPEN 486

RESULT 11  
 E96716  
 Probable RNA-binding protein F23010.17 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: E96716  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Huizar, L.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: E96716  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-427 <STO>  
 A:Cross-references: GB:AE005173; NID:97705097; PIDN:AAF67776.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F23010.17  
 A:Map position: 1

Query Match 10.8%; Score 81; DB 2; Length 427;  
 Best Local Similarity: 23.7%; Pred. No. 3.6;  
 Matches 33; Conservative 26; Mismatches 62; Indels 18; Gaps 5;  
 QY 11 VDOACRAAEFVNIYETMDKRRALTRLYLDKATLI---WNGNAV--GLDALNNFFDT 65  
 Db 7 VPSAQDIAAEFFVQYVYVGLQHPHARRLYVDASVSRPDVTGTMMSTVSVEAINKHLS 66  
 QY 66 --LPSEFQVNMDCQPVHEQATQSTTVLVVTSQVTFKDGKQHFENQNFLLTAQSTPN 123  
 Db 67 CDFENTKFEVLSVDSONSLDG-----IFIMVIGTGTGNORRKFSONFYLARQNT-- 118  
 QY 124 NTWKIATSDCFRFDWSSS 142  
 Db 119 ----LVVLNDMLRYVDQEDS 134

RESULT 12  
 S63382  
 hypothetical protein YNR051c - yeast (Saccharomyces cerevisiae)  
 A:Alternate names: hypothetical protein N3465  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 27-Apr-1996 #sequence\_revision 03-May-1996 #text\_change 19-Apr-2002  
 C:Accession: S63382; S27439; S27440  
 R:Pohl, T.M.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S63346  
 A:Accession: S63382  
 A:Molecule type: DNA  
 A:Residues: 1-515 <POH>  
 A:Cross-references: EMBL:Z71666; NID:91302565; PID:e239594; PID:g1302566; MIPS:YNR051  
 A:Experimental source: strain S288C  
 R:Cusick, M.E.  
 submitted to the EMBL Data Library, March 1992  
 A:Reference number: S27437  
 A:Accession: S27439  
 A:Molecule type: DNA  
 A:Residues: 1-316, T' <CUS>  
 A:Cross-references: EMBL:M88607; NID:g172079; PID:g172081  
 A:Accession: S27440  
 A:Molecule type: DNA  
 A:Residues: 350-441 <CUW>  
 A:Cross-references: EMBL:M88607; NID:g172079; PID:g553136  
 C:Genetics:  
 A:Gene: SGD:BRE5  
 A:Cross-references: SGD:S0005334  
 A:Map position: 14R

Query Match 10.8%; Score 81; DB 2; Length 515;  
 Best Local Similarity: 23.0%; Pred. No. 4.6;  
 Matches 34; Conservative 24; Mismatches 56; Indels 34; Gaps 6;  
 QY 11 VDOACRAAEFVNIYETMDKRRALTRLYLDKATLIWNGN-----AVSG 55  
 Db 5 VQDICFA---FLONIYERMTDFPSKLAIFYASTAELTHYQSKSTNEKDDVLPVTKVTG 61  
 QY 56 LDALNNFFD-----TLPSEFQVNMDCQ-----PVHEQATQSTTVLVVTSQVTFKDGKQH 108

## RESULT 13

## RESULT 15

H84450

A; Map pos

Job time

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:51:02 ; Search time 3.39909 Seconds  
(without alignments)  
1732.709 Million cell updates/sec

Title: us-09-763-902b-7

Perfect score: 749

Sequence: 1 MATSLDFKTYVDQACRAAE.....NTYWKIASDCRFQDWSSS 142

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 749   | 100.0       | 142    | 1  | NXT2_HUMAN  |
| 2          | 586   | 78.2        | 140    | 1  | NXT1_MOUSE  |
| 3          | 584   | 78.0        | 140    | 1  | NXT1_HUMAN  |
| 4          | 310   | 41.4        | 133    | 1  | NXT1_DROME  |
| 5          | 207.5 | 27.7        | 137    | 1  | NXT1_CAEEL  |
| 6          | 124   | 16.6        | 122    | 1  | NFT2_ORISA  |
| 7          | 115   | 15.4        | 127    | 1  | NFT2_HUMAN  |
| 8          | 113   | 15.1        | 123    | 1  | NFT2_SCHPO  |
| 9          | 112.5 | 15.0        | 126    | 1  | NFT2_ARATH  |
| 10         | 108.5 | 14.5        | 124    | 1  | NFT2_NEUCR  |
| 11         | 108   | 14.4        | 127    | 1  | NFT2_XENLA  |
| 12         | 104.5 | 14.0        | 124    | 1  | NFT2_CANAL  |
| 13         | 101.5 | 13.6        | 125    | 1  | NFT2_YEAST  |
| 14         | 99.5  | 13.3        | 133    | 1  | NFT2_CAEEL  |
| 15         | 88.5  | 11.8        | 482    | 1  | G3B2_HUMAN  |
| 16         | 88.5  | 11.8        | 482    | 1  | G3B2_MOUSE  |
| 17         | 86.5  | 11.5        | 643    | 1  | N0S2_ALCEU  |
| 18         | 81    | 10.8        | 515    | 1  | YN8T_YEAST  |
| 19         | 78.5  | 10.5        | 450    | 1  | ACHX_HUMAN  |
| 20         | 77.5  | 10.3        | 910    | 1  | Y068_CAEEL  |
| 21         | 77.5  | 10.3        | 309    | 1  | YDAA_HAEIN  |
| 22         | 77.5  | 10.3        | 579    | 1  | UL25_VZVD   |
| 23         | 76.5  | 10.2        | 434    | 1  | G3BP_SCHPO  |
| 24         | 74    | 9.9         | 287    | 1  | FPG_SYNY3   |
| 25         | 73    | 9.7         | 292    | 1  | MCDB_ACICA  |
| 26         | 72.5  | 9.7         | 403    | 1  | YC6F_ECOLI  |
| 27         | 71.5  | 9.5         | 186    | 1  | Y50L_SYNY3  |
| 28         | 71.5  | 9.5         | 1452   | 1  | VGL2_FIPV   |
| 29         | 71.5  | 9.5         | 1939   | 1  | MYH6_HUMAN  |
| 30         | 71    | 9.5         | 278    | 1  | TCPL_VIECH  |
| 31         | 70.5  | 9.4         | 326    | 1  | H0LB_BUCAI  |
| 32         | 70    | 9.3         | 631    | 1  | VEL_HPV30   |
| 33         | 70    | 9.3         | 644    | 1  | VEL_HPV59   |

34 70 9.3 833 1 PT1A\_ECOLI P32670 escherichia  
35 69 9.2 363 1 VG48\_BPT4 P13339 bacterioph  
36 69 9.2 701 1 LSHR\_BOVIN Q28005 bos tauris  
37 68.5 9.1 466 1 G3BP\_HUMAN Q13283 bos sapien  
38 68.5 9.1 753 1 CRNA\_BACUF Q32321 bacillus th  
39 68 9.1 380 1 VE2\_HPV70 P50773 human papil  
40 68 9.1 581 1 YG25\_ARCFU Q28648 archaeoglob  
41 67.5 9.0 259 1 YMEB\_LACLA Q0151 lactococcus  
42 67.5 9.0 465 1 G3BP\_MOUSE P97855 mus musculu  
43 67.5 9.0 710 1 TOP1\_LACLA Q9C980 lactococcus  
44 67.5 9.0 1935 1 MYH7\_HUMAN P12883 homo sapien  
45 67.5 9.0 1935 1 MYH7\_PIG P79293 sus scrofa

#### ALIGNMENTS

##### RESULT 1

NXT2\_HUMAN STANDARD; PRT; 142 AA.  
AC Q9NFJ8: Q9H8U0; Q9NRL7; Q9Y3M4; Q9Y3M5; Q9N064;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE NTF2-related export protein 2 (p15-2 protein) (DC9) (BM025).  
GN NXT2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=restis;  
RX MEDLINE=20528640; PubMed=11073998;  
RA Herold A., Suyama M., Rodrigues J.P., Braun I.C., Kutay U.,  
RA Carmo-Fonseca C., Bork P., Izaurralde E.;  
RT "TAP (NFX1) belongs to a multigene family of putative RNA export  
RT factors with a conserved modular architecture";  
RL Mol. Cell. Biol. 23:8996-9008(2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Kang Y., Cullen B.;  
RT "p15-2, a homologous protein of p15, interacts with Tap.";  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Li Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.;  
RT "Novel genes expressed in human dendritic cell";  
RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Ovarian carcinoma;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamanoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project";  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Bone marrow;  
RA Zhao M., Song H., Li N., Peng Y., Han Z., Chen Z.;  
RT "A novel gene expressed in human bone marrow";  
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RA Howden P.;  
RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Regulator of protein export for NES-containing proteins.  
CC Also plays a role in mRNA nuclear export.

```

CC CC -!- SUBUNIT: Associates with NXF1, NXF2, NXF3 and NXF5.
CC CC -!- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and
CC CC the cytoplasm.
CC CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; A (shown here) and B; are
CC CC produced by alternative splicing.
CC CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; AJ277591; CAB96371.1; -
CC CC EMBL; AJ278323; CAC01129.1; -
CC CC EMBL; AF246127; AAF78034.1; -
CC CC EMBL; AF201942; AAF86878.1; -
CC CC EMBL; AK023289; BAB14511.1; ALT_INIT.
CC CC EMBL; AF212223; AAF87325.1; ALT_INIT.
CC CC EMBL; AL031387; CAB41301.1; -
CC CC EMBL; AL031387; CAB41302.1; -
CC CC MIM; 300320; -
CC CC InterPro; IPR002075; NTF2.
CC CC Pfam; PF02136; NTF2; 1.
CC CC PROSITE; PS50177; NTF2_DOMAIN; 1.
CC CC Transport; Protein transport; mRNA transport; Nuclear protein;
CC CC Alternative splicing.
CC CC DOMAIN 17 136
CC CC FT VARSPLIC 1 5 NTF2.
CC CC FT SEQUENCE 142 AA; 16228 MW; 009B4E929A6BA6C CRC64;
CC CC
CC CC Query Match 100.0%; Score 749; DB 1; Length 142;
CC CC Best Local Similarity 100.0%; Pred. No. 1.3e-70;
CC CC Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 1 MATSLDKFTYVDQACRAAEFEVNIYETMDKRRRLATRLYLDKATLIWNGNAVSGLDALN 60
CC CC DB 1 MATSLDKFTYVDQACRAAEFEVNIYETMDKRRRLATRLYLDKATLIWNGNAVSGLDALN 60
CC CC
CC CC QY 61 NFDTLPSSEFQVNMDCQPVHEQATQSTVLVWTSQTVKFDGKQHFNFQNLTAQS 120
CC CC DB 1 NFDTLPSSEFQVNMDCQPVHEQATQSTVLVWTSQTVKFDGKQHFNFQNLTAQS 120
CC CC
CC CC QY 61 NFDTLPSSEFQVNMDCQPVHEQATQSTVLVWTSQTVKFDGKQHFNFQNLTAQS 120
CC CC DB 1 NFDTLPSSEFQVNMDCQPVHEQATQSTVLVWTSQTVKFDGKQHFNFQNLTAQS 120
CC CC
CC CC QY 121 TPNTVWKIASDCFRFQDWSS 142
CC CC DB 121 TPNTVWKIASDCFRFQDWSS 142
CC CC
CC CC RESULT 2
CC CC NNTL_MOUSE STANDARD; PRT; 140 AA.
CC CC AC Q9QV9; Q9DIP7;
CC CC DT 16-OCT-2001 (Rel. 40, Created)
CC CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC CC DE NTF2-related export protein 1.
CC CC GN NTF1.
CC CC OS Mus musculus (Mouse).
CC CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC CC OX NCBI_Taxid=10090;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RX MEDLINE-20036817; PubMed-10567585;
CC CC RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;
CC CC RT "Identification of an NTF2-related factor that binds Ran-GTP and
CC CC regulates nuclear protein export.";
CC CC RL Mol. Cell. Biol. 19:8616-8624(1999).
CC CC RN [2]
CC CC RP SEQUENCE FROM N.A.

```

```

RC STRAIN-C57BL/6J; TISSUE=Embryo;
RX MEDLINE-21085660; PubMed-11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Tamaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W.J., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
CC CC Nature 409:685-690(2001).
CC CC -!- FUNCTION: Stimulator of protein export for NES-containing
CC CC proteins. Also plays a role in the nuclear export of U1 snRNA,
CC CC tRNA, and mRNA (By similarity).
CC CC -!- SUBUNIT: Preferentially binds Ran-GTP. Associates with NXF1, NXF2
CC CC and NXF3.
CC CC -!- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and
CC CC the cytoplasm.
CC CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; AF156958; AAD54943.2; -
CC CC EMBL; AK003254; BAB22670.1; -
CC CC MGD; MGI:1929619; Nxf1.
CC CC InterPro; IPR002075; NTF2.
CC CC Pfam; PF02136; NTF2; 1.
CC CC PROSITE; PS50177; NTF2_DOMAIN; 1.
CC CC Transport; Protein transport; mRNA transport; Nuclear protein.
CC CC FT DOMAIN 16 135
CC CC FT CONFLICT 100 100 K -> N (IN REF. 2).
CC CC SQ SEQUENCE 140 AA; 15847 MW; 358EAFDC19FE4594 CRC64;
CC CC
CC CC Query Match 78.2%; Score 586; DB 1; Length 140;
CC CC Best Local Similarity 76.1%; Pred. No. 9.3e-54;
CC CC Matches 105; Conservative 20; Mismatches 13; Indels 0; Gaps 0;
CC CC
CC CC QY 4 SLDFKTYVDQACRAAEFEVNIYETMDKRRRLATRLYLDKATLIWNGNAVSGLDALNFF 63
CC CC DB 3 SVDFKTYVDQACRAAEFEVNIYETMDKRRRLATRLYLDKATLIWNGNAVSGLDALNFF 62
CC CC
CC CC QY 64 DTLPSSEFQVNMDCQPVHEQATQSTVLVWTSQTVKFDGKQHFNFQNLTAQSTPN 123
CC CC DB 63 EMLPSSEFQVNMDCQPVHEQATQSTVLVWTSQTVKFDGKQHFNFQNLTAQSTPN 122
CC CC
CC CC QY 124 NTWKIASDCFRFQDWSS 141
CC CC DB 123 NTWKIASDCFRFQDWSS 140
CC CC
CC CC RESULT 3
CC CC ID NNTL_HUMAN STANDARD; PRT; 140 AA.
CC CC AC Q9UKK6;
CC CC DT 16-OCT-2001 (Rel. 40, Created)
CC CC DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF156957; AAD54942.1; -  
DR EMBL; AL096677; CAC03437.2; -  
DR EMBL; BC000759; AAH00759.1; -  
DR EMBL; BC002687; AAH02687.1; -  
DR EMBL; BC003029; AAH03029.1; -  
DR EMBL; BC003410; AAH03410.1; -  
DR GenBank; HGNC:15913; NTF1.  
DR MIM; 605811; -  
DR InterPro; IPR002075; NTF2.  
DR Pfam; PF02136; NTF2; 1.  
DR PROSITE; PSS0177; NTF2\_DOMAIN; 1.  
KW Transport; Protein transport; Nuclear protein.  
FT DOMAIN 16 135  
SQ SEQUENCE 140 AA; 15847 MW; 358FA86AC3944594 CRC64;

Query Match 78.0%; Score 584; DB 1; Length 140;  
Best Local Similarity 75.4%; Pred. No. 1.5e-53;  
Matches 104; Conservative 22; Mismatches 12; Indels 0; Gaps 0;

QY 4 SLOPKTVDDACRAAEFVNIYETMDKRRALRLYLKATLWNGNAVSGLDALNPF 63  
Db 3 SVDPKTVDDACRAAEFVNIYETMDKRRALRLYLKATLWNGNAVSGLDALNPF 62  
QY 64 DTPSSFFVQVMDLCPVHEQATQSQTVLVTVTSFGVFDGKQHFNFQFLTLTAQSTPN 123  
Db 63 EMLPSSEFQVMDLCPVHEQATQSQTVLVTVTSFGVFDGKQHFNFQFLTLTAQSTPN 122  
QY 124 NTWKIASDCFRQDWSS 141  
Db 123 NTWKIASDCFRQDWAS 140

RESULT 4  
NXTL\_DROME  
ID NXTL\_DROME STANDARD; PRT; 133 AA.  
AC Q9V3H8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE NTF2-related export protein (p15).  
GN NXT1 OR CGI2752.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20036817; PubMed-10567585;  
RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;  
RT "Identification of an NTF2-related factor that binds Ran-GTP and  
RT regulates nuclear protein export.";  
RL Mol. Cell. Biol. 19:8616-8624(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN-Berkeley.  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agayani A., An H.-J., Andrews-Birnko C., Baldwin D.,

15-JUN-2002 (Rel. 41, Last annotation update)  
DE NTF2-related export protein 1 (p15 protein).  
GN NXT1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND FUNCTION.  
RX MEDLINE-20036817; PubMed-10567585;  
RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;  
RT "Identification of an NTF2-related factor that binds Ran-GTP and  
RT regulates nuclear protein export.";  
RL Mol. Cell. Biol. 19:8616-8624(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21638749; PubMed-11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
RA Jones M., Stavridis G., Almeida J.P., Babbage A.K., Bagdole C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond S., Harley J.R., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Huckle E., Hunt A.L., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor J.D.,  
RA Levasalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,  
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,  
RA Stuke C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
RA Svann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA TISSUE=Lung, Placenta, and Uterus;  
RA Strausberg R.;  
RN Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP FUNCTION.  
RX MEDLINE-20307861; PubMed-10848583;  
RA Ossareh-Nazari B., Malsion C., Black B.E., Levesque L., Paschal B.M.,  
RA Dargemont C.;  
RT "RanGTP-binding protein NXT1 facilitates nuclear export of different  
RT classes of RNA in vitro.";  
RL Mol. Cell. Biol. 20:4562-4571(2000).  
RN [5]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [6]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [7]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [8]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [9]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [10]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [11]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [12]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [13]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [14]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D., Hamarskjold M.L.;  
RT "NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of  
RT intron-containing RNA in mammalian cells.";  
RL Mol. Cell. Biol. 21:2545-2554(2001).  
RN [15]  
RP FUNCTION.  
RX MEDLINE-21160285; PubMed-11259602;  
RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
RA Paschal B.M., Rekosh D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cayley S., Dahle C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M., D.L.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weltschbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT The genome sequence of *Drosophila melanogaster*.  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: Stimulator of protein export for NES-containing  
 CC proteins. Also plays a role in the nuclear export of. U1 snRNA,  
 CC tRNA, and mRNA (By similarity).  
 CC -1- SUBUNIT: Preferentially binds Ran-GTP (By similarity). Associates  
 CC with NXF1.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AF156959; AAD54944.1; -  
 CC EMBL; AE003462; AAF47066.1; -  
 CC FlyBase; FBgn0028411; Nxt1.  
 CC InterPro; IPR002075; NTF2.  
 CC Pfam; PF02136; NTF2; 1.  
 CC PROSITE; PS50177; NTF2\_DOMAIN; 1.  
 KW Transport; Protein transport; mRNA transport; Nuclear protein.  
 FT DOMAIN 15 130 NTF2.  
 SQ SEQUENCE 133 AA; 15181 MW; C6E664950AA370AA CRC64;

Query Match 41.4%; Score 310; DB 1; Length 133;  
 Best Local Similarity 42.9%; Pred. No. 3.2e-25;  
 Matches 57; Conservative 24; Mismatches 48; Indels 4; Gaps 1;  
 QY 6 DFKTVQDQACRAAEFVNIYETMDKRRALTRLYLDKATLIWNGVNSGLDALNNFFDT 65  
 Db 4 DLKAKVESCARTADTRLYAYSDVNRQIGRLYLDNLATLSWNGAIGRMIESYFOE 63  
 QY 66 LPSSEFQVNLDCQPVHEQATQSTTVLVVTSCTGVKFDGNKQHFNFQNFLLTAQSTPNNT 125  
 Db 64 LPSSNHLNTLDAQPIVDQASVNLAYLTMASGSVRFADQQLKFOQTFTVTAEE---ND 119  
 QY 126 VKIADSCRFQD 138  
 Db 120 KKWVSDCYRQAE 132

RESULT 5  
 Nxt1\_CAEEL STANDARD; PRT; 137 AA.  
 ID Nxt1\_CAEEL  
 AC Q9U757;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 13-JUN-2002 (Rel. 41, Last annotation update)  
 DE NTF2-related export protein.  
 GN NTF1 OR Y71F9AM.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID:6239;  
 RN [1]  
 RP MEDLINE=20036817; PubMed=10567585;  
 RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;  
 RT "Identification of an NTF2-related factor that binds Ran-GTP and  
 RT regulates nuclear protein export";  
 RL Mol. Cell. Biol. 19:8616-8624(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA Bradshaw-Cordum H., Scott K., Graves T.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Stimulator of protein export for NES-containing  
 CC proteins. Also plays a role in the nuclear export of. U1 snRNA,  
 CC tRNA, and mRNA (By similarity).  
 CC -1- SUBUNIT: Preferentially binds Ran-GTP (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AF156960; AAD54945.1; -  
 CC EMBL; AC024852; AAK66028.1; -  
 CC HSP; P13662; 10UN.  
 CC WormPep; Y71F9AM.5; CE26780.  
 CC InterPro; IPR002075; NTF2.  
 CC Pfam; PF02136; NTF2; 1.  
 CC PROSITE; PS50177; NTF2\_DOMAIN; 1.  
 KW Transport; Protein transport; mRNA transport; Nuclear protein.  
 FT DOMAIN 19 135 NTF2.  
 SQ SEQUENCE 137 AA; 15479 MW; B99ABE3CA3479962 CRC64;

Query Match 27.7%; Score 207.5; DB 1; Length 137;  
 Best Local Similarity 29.5%; Pred. No. 1.4e-14;  
 Matches 41; Conservative 35; Mismatches 58; Indels 5; Gaps 2;  
 QY 1 MATSLDFKTYVDQACRAAEFVNIYETMDKRRALTRLYLDKATLIWNGVNSGLDALN 60  
 Db 3 MKTTQEKKEDEELCESKFKMDVYVDMRKREKIGFLYTVQSVNAGVNGNPINGYDSIC 62  
 QY 61 NTFDLTPSEFQVNLDCQPVHEQATQST-TVLVVTSGTVKFDGNKQHFNFQNFLLTAQ 119  
 Db 63 EFMKALPSTQHDIQSLDAQRLPEGVTDMSGGMLLNAGAVTVDGSKRAFTQTLLGLVE 122  
 QY 120 STPNNTVKIADSCRFQD 138  
 Db 123 ----DGKVKRSDFRYVD 137

RESULT 6  
 NTF2\_ORYSA STANDARD; PRT; 122 AA.  
 ID NTF2\_ORYSA  
 AC Q9XU54;





[illegible]

```

RC STRAIN-cv. Columbia;
RX MEDLINE-21016719; PubMed-11130712;
RA Theologin A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Alkafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Laugin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana."
RL Nature 408:816-820(2000).
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE
CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AC069471; AAG51491.1;
CC HSSP: P13662; IOUN.
CC InterPro: IPR002075; NTF2.
CC Pfam: PF02136; NTF2; 1.
CC PROSITE: PS50177; NTF2_DOMAIN; 1.
CC Transport; Protein transport.
CC DOMAIN 9 123 NTF2.
CC SEQUENCE 126 AA; 14002 MW; 263E06A7A8903E3B CRC64;
CC
CC Query Match 15.08; Score 112.5; DB 1; Length 126;
CC Best Local Similarity 27.38; Pred. No. 8.4e-05;
CC Matches 33; Conservative 25; Mismatches 52; Indels 11; Gaps 5;
CC
CC QY 18 AEEFVNIYETMDKRRRLTRLYLDKATLIWNGNAVSGLDALNFFDLP--SSFEQVNM 75
CC Db :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 10 SKAFVHYHYSTFDTNRVGLAGLYOEASMLTFEGQKIQGVQSVIAKLTSLPQOCKHHIST 69
CC
CC QY 76 LDCQPVHEQATQSTTVLVVTSQVTKFGDKQHF-FNQNFLLTAQSTPNNTVWKIASDCF 134
CC Db :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 70 VDCQP-----SGPASGMLVYFSGNLAGEEHALKFSQMFHL--MPTPGGSFY-VFNDIF 121
CC
CC QY 135 R 135
CC Db 122 R 122
CC
CC RESULT 10
CC NTF2_NEUCR
CC ID NTF2_NEUCR STANDARD; PRT; 124 AA.
CC AC P87102;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Nuclear transport factor 2 (NTF-2).
CC GN NTF-2.
CC OS Neurospora crassa.
CC OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
CC Sordariales; Sordariales; Neurospora.

```

```

OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BD-A / FGSC 1858;
RA Meyer U., Meyer M., Teichel D., Toeken K., Rensing L.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE
CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y13237; CAA73689.1;
CC HSSP: P13662; IOUN.
CC InterPro: IPR002075; NTF2.
CC Pfam: PF02136; NTF2; 1.
CC PROSITE: PS50177; NTF2_DOMAIN; 1.
CC Transport; Protein transport.
CC DOMAIN 8 121 NTF2.
CC SEQUENCE 124 AA; 13726 MW; 8C7AAF927B0476C9 CRC64;
CC
CC Query Match 14.58; Score 108.5; DB 1; Length 124;
CC Best Local Similarity 26.58; Pred. No. 0.00021;
CC Matches 35; Conservative 21; Mismatches 63; Indels 13; Gaps 3;
CC
CC QY 4 SLDFKTYVDQACRAAEFVNIYETMDKRRRLTRLYLDKATLIWNGNAVSGLDALNFF 63
CC Db :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 2 SLDFTA-----IATQFVAHYSTFSDSKRNLAGLYRDNLSMLTFEGQSLGAQGITKL 54
CC
CC QY 64 DTLPSSEFQVNMDCQPVHEQATQSTTVLVVTSQVTKFGDKQHF-FNQNFLLTAQSTPN 123
CC Db :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 55 TSLP---FQVKVKEYGPPDPAQPTATGGIILVTGLIVDDEQRPGLGYQAQFSLSDASQ 111
CC
CC QY 124 NTWKIASDCFR 135
CC Db 112 ---NEVFNDIFK 120
CC
CC RESULT 11
CC NTF2_XENLA
CC ID NTF2_XENLA STANDARD; PRT; 127 AA.
CC AC O42242;
CC DT 15-JUL-1999 (Rel. 38, Created)
CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Nuclear transport factor 2 (NTF-2) (P10).
CC GN NTF2.
CC OS Xenopus laevis (African clawed frog).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
CC Xenopodinae; Xenopus.
CC OX NCBI_TaxID=8355;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RA Pu R.T., Shelsta H.N., Dasso M.;
CC RT "Identification of a p10(NTF2) homologue in Xenopus."
CC RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS.
CC INTERACTS WITH THE NUCLEOPORIN P62 AND WITH RAN. ACTS AS A
CC RELATIVELY LATE STAGE OF NUCLEAR PROTEIN IMPORT, SUBSEQUENT TO THE
CC INITIAL DOCKING OF NUCLEAR IMPORT LIGAND AT THE NUCLEAR ENVELOPE.
CC COULD BE PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT
CC ASSEMBLE AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

```

CC CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC  
CC DR EMBL: AF023911; AAB81276.1; .  
CC DR HSP: P13662; 1A2K.  
CC DR InterPro: IPR002075; NTF2.  
CC DR Pfam: PF021136; NTF2; 1.  
CC DR PROSITE: PS0177; NTF2\_DOMAIN; 1.  
CC DR Transport; Protein transport. NTF2.  
CC KW DOMAIN 10 121  
CC FT SEQUENCE 127 AA; 14477 MW; 97110D6828268259 CRC64;  
CC SQ  
Query Match 14.4%; Score 108; DB 1; Length 127;  
Best Local Similarity 27.4%; Pred. No. 0.00025;  
Matches 32; Conservative 16; Mismatches 57; Indels 12; Gaps 3;  
CC  
CC 21 FVNIYEMDKRRRLRLYLKATLWNGAVSGLDANNFFTLPSSEFQ--VNMLDC 78  
CC DB 14 FIOQYOTFADRTQLAVIYTDASCLTWEGQYHGKAAIVEKLSLLPFQKHQHSITSDH 73  
CC  
CC 79 QPVHEQATQSTTVLWTSVGVKDGNGKHFNFQNLITAQSTPNNTVWKIASDCFR 135  
CC DB 74 QPTPD-----SCIISWVGQKADDDPIMGFHQVFL--KNQIDAWCTNDMFR 120  
CC  
CC  
CC RESULT 12  
CC NTF2\_CANAL STANDARD; PRT; 124 AA.  
CC ID NTF2\_CANAL  
CC AC Q9P926;  
CC DT 09-OCT-2001 (Rel. 40, Created)  
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
CC DE Nuclear transport factor 2 (NTF-2).  
CC GN NTF2.  
CC OS Candida albicans (Yeast).  
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
CC OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
CC OX NCBI\_TaxID=5476;  
CC [1]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN=ATCC 26555;  
CC RA Blanchin-Roland S., Cordero-Otero R., Gallardin C., Herrero A.B.,  
CC RA Dominguez A.;  
CC RT Sequence of Candida albicans and Yarrowia lipolytica homolog of the  
CC RT Saccharomyces cerevisiae NTF2 gene.  
CC RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE  
CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE  
CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC  
CC DR EMBL: AF145758; AAF66701.1; .  
CC DR HSP: P13662; 1OUN.  
CC DR InterPro: IPR002075; NTF2.  
CC DR Pfam: PF021136; NTF2; 1.  
CC DR PROSITE: PS0177; NTF2\_DOMAIN; 1.  
CC KW Transport; Protein transport.  
CC

CC FT DOMAIN 8 121 NTF2.  
CC SQ SEQUENCE 124 AA; 14193 MW; B799D1AB828A049 CRC64;  
CC  
CC Query Match 14.0%; Score 104.5; DB 1; Length 124;  
CC Best Local Similarity 28.1%; Pred. No. 0.00055;  
CC Matches 39; Conservative 20; Mismatches 53; Indels 27; Gaps 7;  
CC  
CC 4 SLDFKTYVDQACRAAEFVNIYETMDKRRRLRLYLKATLWNGAVSGLDANNFF 63  
CC DB 2 VDFNA-----VATEFCNFYFNQFSDRSQSLGNLYRNESMLTFE---TSQLOGARDIV 51  
CC  
CC 64 DTLPSSEFQ-----VNMLDCQPVHEQATQSTTVLWTSVGVKDGNGKHFNFQNLITAQSTPNNTVWKIASDCFR 135  
CC DB 52 EKLASLPFQKVAHRISTLDAQP-----ASANGDILVMVTGELLIDEQNAQR-YSQVFEHL 105  
CC  
CC 117 TAOSTPNNTVWKIASDCFR 135  
CC DB 106 ----IPDNGSYVVFNDIFR 120  
CC  
CC  
CC RESULT 13  
CC NTF2\_YEAST STANDARD; PRT; 125 AA.  
CC ID NTF2\_YEAST  
CC AC P33331;  
CC DT 01-FEB-1994 (Rel. 28, Created)  
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)  
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
CC DE Nuclear transport factor 2 (NTF-2) (Nuclear transport factor P10).  
CC GN NTF2 OR YER009W.  
CC OS Saccharomyces cerevisiae (Baker's yeast).  
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
CC OC Saccharomycetales; Saccharomycetales; Saccharomycetes.  
CC OX NCBI\_TaxID=4932;  
CC [1]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN=S288c / AB972;  
CC RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,  
CC RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,  
CC RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkle-Smith S.,  
CC RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,  
CC RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,  
CC RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogen T., Smith V.,  
CC RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;  
CC RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC [2]  
CC RP SEQUENCE OF 1-84 FROM N.A.  
CC RA Haarer B.K., Petzold A.S., Brown S.S.;  
CC RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.  
CC [3]  
CC RP CHARACTERIZATION.  
CC RX MEDLINE=96178681; Pubmed=8600522;  
CC RT Nehrbass U., Blobel G.;  
CC RT Role of the nuclear transport factor p10 in nuclear import.;  
CC RL Science 272:120-122(1996).  
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS...  
CC INTERACTS WITH VARIOUS NUCLEOPORINS AND WITH RAN-GDP. COULD BE  
CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE  
CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT. IN VITRO, THE NTF2-RAN-  
CC GDP ASSOCIATION, IN THE PRESENCE OF GTP, TRIGGERS DISSOCIATION OF  
CC THE KARYOPHERIN ALPHA-BETA COMPLEX, ALLOWING NUCLEAR TRANSLLOCATION  
CC OF KARYOPHERIN ALPHA AND THE NLS SUBSTRATE.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC  
CC DR EMBL: U18778; AAB64542.1; .

DR EMBL: L22204; AAB49379.1;  
 DR PIR: S41793; S41793  
 DR HSP: P13662; IOUN  
 DR SGD: S000811; NTF2  
 DR InterPro: IPR002075; NTF2  
 DR Pfam: PF02136; NTF2; 1  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1  
 KW Transport; Protein transport.  
 FT DOMAIN 8 121 NTF2  
 SQ SEQUENCE 125 AA; 14453 MW; 566080C0E634714 CRC64;

Query Match 13.6%; Score 101.5; DB 1; Length 125;  
 Best Local Similarity 27.98; Pred. No. 0.0011;  
 Matches 38; Conservative 19; Mismatches 58; Indels 21; Gaps 6;

QY 4 SLDPKTYVDQACRAAEFVNIYETMDKRRRLTRLYLDKATLIWNGNAVSGLDALNFF 63  
 Db 2 SLDPNT-----LAQNTQFYQNFDFDRSOLGNLYNESMLFETSOLOKADIVEKL 54  
 QY 64 DTLPSSEFQ--VNMDCQPVHEQATQSTTVLVVTSQVGFED--GNKHQFFNQNFLLTAQ 119  
 Db 55 VSLPFQVQVHRITTLDAQPSPNG-----DVLVMTGLLLIDEQNPQR-ESQVPHL--- 105  
 QY 120 STPNNTVWKIASDCFR 135  
 Db 106 -IPGNSYVVFENDIFR 120

RESULT 14  
 ID NTF2\_CAEL STANDARD; PRT; 133 AA.  
 AC Q21735;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable nuclear transport factor 2 (NTF-2).  
 GN R05D11.3  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA White S.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE  
 CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE  
 CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: 275546; CAA99890.1;  
 DR HSP: P13662; IOUN.  
 DR WormPep: R05D11.3; CE06238.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2; 1  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1  
 KW Transport; Protein transport.  
 FT DOMAIN 10 128 NTF2  
 SQ SEQUENCE 133 AA; 15116 MW; B283AA646D59C7F2 CRC64;

Query Match 13.38; Score 99.5; DB 1; Length 133;  
 Best Local Similarity 26.8%; Pred. No. 0.002;  
 Matches 34; Conservative 27; Mismatches 47; Indels 19; Gaps 6;

QY 18 ABEFNIYYETMD-----KRRRLTRLY-LDKATLIWNGNAVSGLDALNFFDTLPSSSEF 71  
 Db 11 AKAFIOHYYSKFDVGDGKSRAGLSLDYDPENSYTFEQQAQKGRDGLQKTTLTGFTKI 70  
 QY 72 Q--VNMDCQPVHEQATQSTTVLVVTSQVGFEDGNKHQFFNQNFLLTAQSTPNNT-VWK 128  
 Db 71 QRAITVIDSQPLDGSIQ-----VAVLGQLKTDDEDPINPFQSVFIL----RPNNGGSYF 120  
 QY 129 IASDCFR 135  
 Db 121 IGNEIFR 127

RESULT 15  
 ID G3B2\_HUMAN STANDARD; PRT; 482 AA.  
 AC Q9UN86; O75149; O60606; Q9UPAL;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ras-GTPase-activating protein binding protein 2 (GAP SH3-domain  
 DE binding protein 2) (G3BP-2).  
 GN G3BP2 OR KIAA0660.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS A AND B).  
 RC TISSUE-Brain;  
 RA Kennedy D., Mattick J.S.;  
 RT "Characterisation and chromosomal location of G3BP-1 and G3BP-2a/b,  
 RT members of a novel SH3 domain-binding and RNA-binding protein family  
 RT implicated in signal transduction".  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM-A).  
 RC TISSUE-Brain;  
 RX MEDLINE=98403880; PubMed=9734811;  
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. X.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:169-176(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE-Brain;  
 RA Guitard E.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE-Brain;  
 RA Kennedy D., Ru K., Mattick J.S.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE-B-cell;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP INTERACTION WITH IKAPPABALPHA.  
 RX MEDLINE=20549669; PubMed=10969074;  
 RA Prigent M., Barlat I., Langen H., Dargemont C.;  
 RT "Ikappabalpha and Ikappabalpha/NF-kappa B complexes are retained in  
 RT the cytoplasm through interaction with a novel partner, RasGAP  
 RT SH3-binding protein 2.";  
 RL J. Biol. Chem. 275:36441-36449(2000).  
 CC -1- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA  
 CC TRANSPORT (POTENTIAL).  
 CC -1- SUBUNIT: BINDS TO THE N-TERMINAL DOMAIN OF IKAPPABALPHA.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

Wed Mar 5 08:37:56 2003

```
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF145284; AAD51932.1; .
CC EMBL; AB014560; BAA31635.1; .
CC EMBL; AF051311; AAC15705.1; .
CC EMBL; AF053535; AAC95922.1; .
CC EMBL; BC011731; AAH11731.1; .
CC HSSP; P09651; 1HA1.
CC InterPro; IPR002075; NTF2.
CC InterPro; IPR000504; RNA_rec_mot.
CC Pfam; PF00076; rim; 1.
CC Pfam; PF02136; NTF2; 1.
CC SMART; SM00360; RRM; 1.
CC PROSITE; PS00177; NTF2_DOMAIN; 1.
CC PROSITE; PS00102; RRM; 1.
CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
CC Transport; RNA-binding; Alternative splicing.
KW DOMAIN 11 133 NTF2.
FT DOMAIN 331 409 RNA-BINDING (RRM).
FT DOMAIN 114 223 GLU-RICH.
FT DOMAIN 419 479 GLY-RICH.
FT VARSPPLIC 243 275 MISSING (IN ISOFORM B).
FT CONFLICT 267 267 S -> P (IN REF. 2).
FT CONFLICT 359 359 E -> V (IN REF. 3).
FT CONFLICT 460 460 M -> I (IN REF. 3).
SQ SEQUENCE 482 AA; 54111 MW; 2545C6A3F1AAE218; CRC64;

Query Match 11.8%; Score 88.5; DB 1; Length 482;
Best Local Similarity 27.2%; Pred.No. 0.14; Indels 9; Gaps 4;
Matches 34; Conservative 19; Mismatches 63;

Qy 20 EFVNIYYETWDRRRALTRYLDKATLIWGNVSG-----LDALNNFFDTLPSSFEQVN 74
Db 14 EFVQYITLLNKAPYELHRYFGRNRSYVHGVDASGRQEAIVYGQNDIHHKVLNLFSEC 73
Qy 75 MLDQCPVHEQATQQTIVLVVTSCTVKFDGNKQHFNFQNLITAQ-STPNNTVWKIASDC 133
Db 74 HTKIHVDAAHTLSD-GVVVQVMGILLNSGQPERKFMQTFVLAPEGSVPNK--FYVHNDM 130
Qy 134 FRFQD 138
Db 131 FRYED 135

Search completed: March 4, 2003, 15:00:02.
Job time : 5.39909 secs
```

| Result No. | Score | Query |      | Length | DB     | ID     | Description |
|------------|-------|-------|------|--------|--------|--------|-------------|
|            |       | Match | %    |        |        |        |             |
| 1          | 167   | 22.3  | 16.0 | 3      | Q9HDY6 | Q9HDY6 | schizosacch |
| 2          | 163.5 | 21.8  | 16.7 | 5      | Q8T2G7 | Q8T2G7 | dictyosteli |
| 3          | 150   | 20.0  | 12.7 | 5      | Q8T2D5 | Q8T2D5 | dictyosteli |
| 4          | 127.5 | 17.0  | 12.5 | 3      | Q96VNS | Q96VNS | emericeella |
| 5          | 121.5 | 16.2  | 12.3 | 3      | Q9P8H0 | Q9P8H0 | yarrowia li |
| 6          | 120   | 16.0  | 14.6 | 10     | Q8RU32 | Q8RU32 | oryza sativ |
| 7          | 113   | 15.1  | 68.8 | 5      | Q917J5 | Q917J5 | drosohilla  |
| 8          | 113   | 15.1  | 69.0 | 5      | Q9NH72 | Q9NH72 | drosohilla  |
| 9          | 113   | 15.1  | 69.0 | 5      | Q9VF74 | Q9VF74 | drosohilla  |
| 10         | 107.5 | 14.4  | 52.2 | 10     | Q9LBY0 | Q9LBY0 | arabidopsi  |
| 11         | 105   | 14.0  | 122  | 10     | Q9FYK4 | Q9FYK4 | arabidopsi  |
| 12         | 102.5 | 13.7  | 130  | 5      | Q9VJ85 | Q9VJ85 | drosohilla  |
| 13         | 96    | 12.8  | 94.6 | 10     | Q9SRV5 | Q9SRV5 | arabidopsi  |
| 14         | 94    | 12.6  | 537  | 5      | Q21351 | Q21351 | caenorhabd  |
| 15         | 90.5  | 12.1  | 428  | 10     | Q9LWX6 | Q9LWX6 | arabidopsi  |
| 16         | 87.5  | 11.7  | 460  | 10     | Q9FWE2 | Q9FWE2 | arabidopsi  |

```
ID Q8T2G7 PRELIMINARY; PRT; 167 AA.
AC Q8T2G7;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 19.0 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115607; AAL92371.1;
KW Hypothetical protein.
SQ SEQUENCE 167 AA; 19039 MW; E16E4A5F4DA00BAA CRC64;

Query Match 21.8%; Score 163.5; DB 5; Length 167;
Best Local Similarity 31.0%; Pred. No. 3.8e-09;
Matches 40; Conservative 18; Mismatches 68; Indels 3; Gaps 3;

QY 8 KTVDAQRAAEFV-NIYETMDKRRRLRLYLKATLIWNGNAVSGLDALNFFDTL 66
DB 38 KKVETITPRAEVFIKEFYIPKYDSSRADLIGLYKDHVSINWNGTECKPGEHIGKLLAEI 97
QY 67 PSSEFOVNMDCOPVHEQATQSTTVLVTSQVTKFDGNKQHFNFQNLFLTAQSTPNNTV 126
DB 98 PNVSHVVEITDAQVPESD-DKENPNLITATGKVTYKTSQHQFHQTLFLVLDKPT-NSNL 155
QY 127 WKIASDCFR 135
DB 156 FYLSYDCIR 164

RESULT 3
Q8T2D5 PRELIMINARY; PRT; 127 AA.
ID Q8T2D5;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative nuclear transport factor similar to nuclear transport factor
DE 2 (NTF2) from.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115613; AAM10776.1;
SQ SEQUENCE 127 AA; 14317 MW; 2D45B1B20021B298 CRC64;

Query Match 20.0%; Score 150; DB 5; Length 127;
Best Local Similarity 29.4%; Pred. No. 6.8e-08;
Matches 37; Conservative 22; Mismatches 55; Indels 12; Gaps 4;

QY 13 QACRAAEFVNIYETMDKRRRLRLYLKATLIWNGNAVSGLDALNFFDTLP--SSE 70
DB 7 QVGVGKQVEHYGIFDSNRAGLQIQOQQTLLTWEGFLSGDAIYKHIVELPFQQTN 66
QY 71 FOVNMDCOPVHEQATQSTTVLVTSQVTKFDGNKQHFNFQ-NFLLLTAQSTPNNTVVKI 129
DB 67 RKINSIDCOQTY-----QPGIMITVTGLIIDGEAK---NQLKFEVFNLSNNGSFL 117
```

```
QY 130 ASDCFR 135
DB 118 INDFFR 123

RESULT 4
Q96VN3 PRELIMINARY; PRT; 125 AA.
ID Q96VN3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Nuclear transport factor 2.
GN NTF2.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA Espeso E.A., Penalva M.A.;
RT "Nuclear transport factor 2 from Aspergillus nidulans."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY038983; AAK71467.1;
DR InterPro; IPR002075; NTF2.
DR Pfam; PF02136; NTF2; 1.
SQ SEQUENCE 125 AA; 13871 MW; 8FC18BDFC597E2F2 CRC64;

Query Match 17.0%; Score 127.5; DB 3; Length 125;
Best Local Similarity 29.8%; Pred. No. 1.4e-05;
Matches 36; Conservative 21; Mismatches 53; Indels 11; Gaps 4;

QY 18 ABEEFVNIYETMDKRRRLRLYLKATLIWNGNAVSGLDALNFFDTLP--SSEFOVNM 75
DB 8 AQQVTFYIYQTFDGNRAGLAPLYRHSMLTFTFSAIQGVAGIIEKLTSLPFQKVQHVST 67
QY 76 LDCQPVHEQATQSTTVLVTSQVTKFDGNKQHFNFQNLFLTAQSTPNNT-VKMIASDCF 134
DB 68 LDAQPSGEGH-----GILVLTGALLVDEKN---PMNTQTQFLMPDAGSYFVLNDVF 119
QY 135 R 135
DB 120 R 120

RESULT 5
Q9P8H0 PRELIMINARY; PRT; 123 AA.
ID Q9P8H0;
AC Q9P8H0;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Rpl1.
GN RPL1.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RA Cordero Otero R.R., Lepingle A., Gaillardin C.;
RT "RPL1 gene from Yarrowia lipolytica."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Rojas Quijano R., Lepingle A., Gaillardin C.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF260231; AAF70316.1;
DR HSP; P13662; 10UN.
DR InterPro; IPR002075; NTF2.
DR Pfam; PF02136; NTF2; 1.
SQ SEQUENCE 123 AA; 13696 MW; 3E6CB95CF0058251 CRC64;
```

```

Query Match 16.2% Score 121.5; DB 3; Length 123;
Best Local Similarity 27.6% Pred. No. 5.9e-05;
Matches 37; Conservative 23; Mismatches 57; Indels 17; Gaps 4;

QY 4 SLDPKTYVDQACRAAEFVNIYETMDKRRRLRLYLDKATLIWNGNAVSGLDALNFF 63
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 SVDFNT-----LAKQCFEYQTFDTDSGLNLYRDHSMLETFQHQGAQIVK 54

QY 64 DTLPSSE--FQVNLDCQPVHEQATQSTVTLVVTSGTVKFDGNKHQFNFLTAOST 121
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 VGLPFGVRRHKISDIAQPSAOGD-----VIVLTGELCVGDGNPLPYGVQVHL----I 106

QY 122 PNTVWVKIASDCFR 135
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107 PDGSSYVFNDFR 120

RESULT 6
ID Q8RU32 PRELIMINARY; PRT; 146 AA.
AC Q8RU32;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative nuclear transport factor 2.
GN P0415A04.24.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
 clone:P0415A04."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003345; BAB90110.1;
SQ SEQUENCE 146 AA; 15796 MW; 496BE9A02E1E9670 CRC64;

Query Match 16.0% Score 120; DB 10; Length 146;
Best Local Similarity 30.5% Pred. No. 0.0001;
Matches 39; Conservative 18; Mismatches 59; Indels 12; Gaps 5;

QY 12 DQACRAAEFVNIYETMDKRRRLRLYLDKATLIWNGNAVSGLDALNFFDTLPSSSEF 71
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23 DQCDVWARAFVEYIYQFTDTRALALYGGTSMLSFSGHNVAGAEIGRKLGLPFEQC 82

QY 72 Q--VNMLDCQPVHEQATQSTVTLVVTSGTVKFDGNKHQF--FNONFLTAOSTPNT--VW 127
 : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 RHAVCTVDCP-----TPSPFGGILVFSGNLQAGAEHQLRFSQMF-----QLVPNEQGSF 134

QY 128 KIASCDFR 135
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 FVQNDIFR 142

RESULT 7
ID Q917J5 PRELIMINARY; PRT; 688 AA.
AC Q917J5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Rln protein.
GN RIN OR CG9412.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong E.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003701; AAG22151.1;
DR FlyBase; FBgn0015778; rln.
DR InterPro; IPR002075; NTF2.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF02136; NTF2; 1.
DR Pfam; PF00076; rlm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS00102; RRM; 1.
SQ SEQUENCE 688 AA; 74713 MW; 9E329949E7248C19 CRC64;

Query Match 15.1% Score 113; DB 5; Length 688;
Best Local Similarity 31.5% Pred. No. 0.004;
Matches 39; Conservative 14; Mismatches 55; Indels 16; Gaps 5;

QY 20 EFNVIYETMDKRRRLRLYLDKATLI--WNGNAVSGLDALNFFDTLPSSSEFQVNMDC 78
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17 EFNVIYETMDKRRRLRLYLDKATLI--WNGNAVSGLDALNFFDTLPSSSEFQVNMDC 69

QY 79 ---QPVEQATQSTVTLVVTSGTVKFDGNKHQFNFLTAOSTPNTVWKTASDCF 134
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 HAKISQVDAQATLNGVGVVQVT--GELSNDGQPMRRRTQTFVLAASPKK---YYVNDIF 125

QY 135 RFQD 138
 |||
Db 126 RVQD 129

RESULT 8
Q9NH72 PRELIMINARY; PRT; 690 AA.
ID Q9NH72
AC Q9NH72;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

```



DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Rasputin.  
 GN RIN OR CG9412.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20191744; PubMed=10725247;  
 RA Pazman C., Mayes C.A., Fanto M., Haynes S.R., Mlodzik M.;  
 RT "Rasputin, the Drosophila homolog of the RasGAP SH3 binding protein,  
 RT functions in Ras and Rho mediated signaling."  
 RL Development 127:1715-1725(2000).  
 DR EMBL; AF310331; AAF68949.1;  
 DR FlyBase; FBgn0015778; rin.  
 DR InterPro; IPR002075; NTF2.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF02136; NTF2; 1.  
 DR Pfam; PF00076; rin; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 SQ SEQUENCE 690 AA; 74985 MW; 40C5BD7E9F5FA253 CRC64;  
  
 Query Match 15.1%; Score 113; DB 5; Length 690;  
 Best Local Similarity 31.5%; Pred. No. 0.004;  
 Matches 39; Conservative 14; Mismatches 55; Indels 16; Gaps 5;  
  
 QY 20 EFVNIYETMDKRRRLRLYLDKATLI-WNGNAVSGLDALNNFFDTLPSSSEFQVNMDC 78  
 DB 19 EFVQYITLLNKAPNHLRFYNNHSSYIHGSKLVVGOREIHNRIQ-----QLNFNDC 71  
 QY 79 ----OPVHEQATQSOTTVLVVTSGVFKDGNKQHFNFQNLTAQSTPNTNTWKIASDCF 134  
 DB 72 HAKISQVDAQTLGNGVVQVVT-GELSDNGQPMRRFTQTFVLAQSPKK---YYVHNDIF 127  
  
 QY 135 RFQD 138  
 DB 128 RYQD 131  
  
 RESULT 9  
 QSVFT4 ID QSVFT4 PRELIMINARY; PRT; 690 AA.  
 AC QSVFT4;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE CG9412 protein (LD31194P).  
 GN RIN OR CG9412.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=107311132;  
 RA Adams M.D., Celnikner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scher S.E., Li P., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake J.P., Champ M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Prannkoc C., Baldwin D.,  
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Fertaz C., Ferrera S., Fleischmann W.,  
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K., Nixon K., Nusskern D.R., Pacleeb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nunoo J., Pacleeb J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celnikner S.;  
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AE030701; AAF54965.1;  
 DR EMBL; AY058617; AAL13846.1;  
 DR FlyBase; FBgn0015778; rin.  
 DR InterPro; IPR002075; NTF2.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF02136; NTF2; 1.  
 DR Pfam; PF00076; rin; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 SQ SEQUENCE 690 AA; 74943 MW; 40C5AC6F9F4EB353 CRC64;  
  
 Query Match 15.1%; Score 113; DB 5; Length 690;  
 Best Local Similarity 31.5%; Pred. No. 0.004;  
 Matches 39; Conservative 14; Mismatches 55; Indels 16; Gaps 5;  
  
 QY 20 EFVNIYETMDKRRRLRLYLDKATLI-WNGNAVSGLDALNNFFDTLPSSSEFQVNMDC 78  
 DB 19 EFVQYITLLNKAPNHLRFYNNHSSYIHGSKLVVGOREIHNRIQ-----QLNFNDC 71  
 QY 79 ----OPVHEQATQSOTTVLVVTSGVFKDGNKQHFNFQNLTAQSTPNTNTWKIASDCF 134  
 DB 72 HAKISQVDAQTLGNGVVQVVT-GELSDNGQPMRRFTQTFVLAQSPKK---YYVHNDIF 127  
  
 QY 135 RFQD 138  
 DB 128 RYQD 131  
  
 RESULT 10  
 QSVFT4 ID QSVFT4 PRELIMINARY; PRT; 522 AA.  
 AC QSVFT4;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE T23J18.22.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;







NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-08525-24

Query Match 9.5%; Score 71.5; DB 5; Length 748;  
Best Local Similarity 27.0%; Pred. No. 6.3;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

QY 36 LRLYLKATLIWNGNAVSGLDALNNFFDTLPSSFEQVNMLD-----CQPVHEQATQS 88  
DB 211 VLLYSRSSTATWEYSAAAYQGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263  
QY 89 QTTVLVVTSGTVKFDGNKQHFNFQNLFLTAQST-----PN-----NTVWKIAS 131  
DB 264 ATNWFAPTSGGYIPDG---FSFNWFLNTNSTFVSGRFTVNTQPLLINCLWPVPS 315

RESULT 6  
PCT-US91-08525-28  
Sequence 28, Application PC/TUS9108525  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
TITLE OF INVENTION: Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/08525  
FILING DATE: 19911114  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-08525-28

Query Match 9.5%; Score 71.5; DB 5; Length 748;

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/392,459  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-392-459-34

Query Match 9.5%; Score 71.5; DB 4; Length 748;  
Best Local Similarity 27.0%; Pred. No. 6.3;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

QY 36 LRLYLKATLIWNGNAVSGLDALNNFFDTLPSSFEQVNMLD-----CQPVHEQATQS 88  
DB 211 VLLYSRSSTATWEYSAAAYQGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263  
QY 89 QTTVLVVTSGTVKFDGNKQHFNFQNLFLTAQST-----PN-----NTVWKIAS 131  
DB 264 ATNWFAPTSGGYIPDG---FSFNWFLNTNSTFVSGRFTVNTQPLLINCLWPVPS 315

RESULT 5  
PCT-US91-08525-24  
Sequence 24, Application PC/TUS9108525  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
TITLE OF INVENTION: Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/08525  
FILING DATE: 19911114  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:

Best Local Similarity 27.0%; Pred. No. 6.3;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;  
QY 36 LTRLYLKATLIWGNVAVSGLDALNNFFDTLPSSFEQVNMLD-----CQPVHEQATQS 88  
Db 211 VTLXSRSTATWEYSAAAYQGVSNF-----TYIKLNTNGLKTYELCED-YEHCTGY 263  
QY 89 OTTVLVVTSQVTKFDGNKHQHFNFQNLFLTAOST-----PN-----NTVWKIAS 131  
Db 264 ATNVFAPTSGGYIPDG---FSFNWFLNLSSTFVSGRFTVNOPLLINCLWPVPS 315

## RESULT 7

PCT-US91-08525-34  
Sequence 34, Application PC/TUS9108525  
GENERAL INFORMATION:

APPLICANT: Smithline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
TITLE OF INVENTION: Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/08525  
FILING DATE: 19911114

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:

NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 748 amino acids

TYPE: AMINO ACID

TOPOLOGY: unknown

MOLECULE TYPE: peptide

PCT-US91-08525-34

## Query Match

Best Local Similarity 9.5%; Score 71.5; DB 5; Length 748;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

QY 36 LTRLYLKATLIWGNVAVSGLDALNNFFDTLPSSFEQVNMLD-----CQPVHEQATQS 88  
Db 211 VTLXSRSTATWEYSAAAYQGVSNF-----TYIKLNTNGLKTYELCED-YEHCTGY 263  
QY 89 OTTVLVVTSQVTKFDGNKHQHFNFQNLFLTAOST-----PN-----NTVWKIAS 131  
Db 264 ATNVFAPTSGGYIPDG---FSFNWFLNLSSTFVSGRFTVNOPLLINCLWPVPS 315

## RESULT 8

PCT-US93-04384-4

Sequence 4, Application PC/TUS9304384

GENERAL INFORMATION:  
APPLICANT: Miller, Timothy J.  
APPLICANT: Klepper, Sharon  
APPLICANT: Reed, Albert Paul  
APPLICANT: Jones, Elaine V.  
TITLE OF INVENTION: Compositions and Methods for Vaccination  
TITLE OF INVENTION: Against Coronaviruses  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporation - Corporate  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04384  
FILING DATE: 19930507

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/882,171  
FILING DATE: 08-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990

ATTORNEY/AGENT INFORMATION:

NAME: Schreck, Patricia A.  
REGISTRATION NUMBER: 33,777

REFERENCE/DOCKET NUMBER: SBC H85009-1  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 748 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US93-04384-4

## Query Match

Best Local Similarity 9.5%; Score 71.5; DB 5; Length 748;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

QY 36 LTRLYLKATLIWGNVAVSGLDALNNFFDTLPSSFEQVNMLD-----CQPVHEQATQS 88  
Db 211 VTLXSRSTATWEYSAAAYQGVSNF-----TYIKLNTNGLKTYELCED-YEHCTGY 263  
QY 89 OTTVLVVTSQVTKFDGNKHQHFNFQNLFLTAOST-----PN-----NTVWKIAS 131  
Db 264 ATNVFAPTSGGYIPDG---FSFNWFLNLSSTFVSGRFTVNOPLLINCLWPVPS 315

## RESULT 9

PCT-US93-04384-6

Sequence 6, Application PC/TUS9304384

GENERAL INFORMATION:

APPLICANT: Miller, Timothy J.  
APPLICANT: Klepper, Sharon

APPLICANT: Reed, Albert Paul  
APPLICANT: Jones, Elaine V.

TITLE OF INVENTION: Compositions and Methods for Vaccination  
TITLE OF INVENTION: Against Coronaviruses

NUMBER OF SEQUENCES: 48

RESULT 11  
 US-08-392-459-26  
 ; Sequence 26, Application US/08392459  
 ; Patent No. 6280974  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 ; APPLICANT: SmithKline Beecham, Corporation  
 ;  
 ; TITLE OF INVENTION: Recombinant Feline Coronavirus S  
 ;  
 ; TITLE OF INVENTION: Proteins  
 ;  
 ; NUMBER OF SEQUENCES: 54  
 ;  
 ; CORRESPONDENCE ADDRESS:  
 ;  
 ; ADDRESSER: SmithKline Beecham Corporation  
 ;  
 ; STREET: 709 Swedeland Road  
 ;  
 ; CITY: King of Prussia  
 ;  
 ; STATE: PA  
 ;  
 ; COUNTRY: USA  
 ;  
 ; ZIP: 19406-2799  
 ;  
 ; COMPUTER READABLE FORM:  
 ;  
 ; MEDIUM TYPE: Floppy disk  
 ;  
 ; COMPUTER: IBM PC compatible  
 ;  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ;  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ;  
 ; CURRENT APPLICATION DATA:  
 ;  
 ; APPLICATION NUMBER: US/08/392,459  
 ;  
 ; FILING DATE:  
 ;  
 ; CLASSIFICATION: 435  
 ;  
 ; PRIOR APPLICATION DATA:  
 ;  
 ; APPLICATION NUMBER: US 07/698,927  
 ;  
 ; FILING DATE: 13-MAY-1991  
 ;  
 ; PRIOR APPLICATION DATA:  
 ;  
 ; APPLICATION NUMBER: US 07/613,066  
 ;  
 ; FILING DATE: 14-NOV-1990  
 ;  
 ; ATTORNEY/AGENT INFORMATION:  
 ;  
 ; NAME: King, William T.

REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-392-459-26

Query Match 9.5%; Score 71.5; DB 4; Length 1454;  
Best Local Similarity 27.0%; Pred. No. 17;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;  
QY 36 LTRLYDKATLIWGNVAGSLDANNFFDTPSSSEFQVNMLD-----CQPVHEQATQS 88  
DB 211 VTLVSRSTATWEYSAAYAGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263  
QY 89 QTTVLVVTSGTVKFDGNKQHFNFQNFLLTAQST-----PN-----NTVWKIAS 131  
DB 264 ATNVFAPTSGGYIPDG---FSFNNWFLNTSSTFVSGRFTVNOPLLINCLWPVPS 315

## RESULT 12

PCT-US91-08525-22  
Sequence 22, Application PC/TUS9108525  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
TITLE OF INVENTION: Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/08525  
FILING DATE: 19911114  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-08525-22

Query Match 9.5%; Score 71.5; DB 5; Length 1454;  
Best Local Similarity 27.0%; Pred. No. 17;

Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;  
QY 36 LTRLYDKATLIWGNVAGSLDANNFFDTPSSSEFQVNMLD-----CQPVHEQATQS 88  
DB 211 VTLVSRSTATWEYSAAYAGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263  
QY 89 QTTVLVVTSGTVKFDGNKQHFNFQNFLLTAQST-----PN-----NTVWKIAS 131  
DB 264 ATNVFAPTSGGYIPDG---FSFNNWFLNTSSTFVSGRFTVNOPLLINCLWPVPS 315

## RESULT 13

PCT-US91-08525-26  
Sequence 26, Application PC/TUS9108525  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
TITLE OF INVENTION: Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/08525  
FILING DATE: 19911114  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-08525-26

Query Match 9.5%; Score 71.5; DB 5; Length 1454;  
Best Local Similarity 27.0%; Pred. No. 17;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

QY 36 LTRLYDKATLIWGNVAGSLDANNFFDTPSSSEFQVNMLD-----CQPVHEQATQS 88  
DB 211 VTLVSRSTATWEYSAAYAGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263  
QY 89 QTTVLVVTSGTVKFDGNKQHFNFQNFLLTAQST-----PN-----NTVWKIAS 131  
DB 264 ATNVFAPTSGGYIPDG---FSFNNWFLNTSSTFVSGRFTVNOPLLINCLWPVPS 315

## RESULT 14

PCT-US93-04384-2  
Sequence 2, Application PC/TUS9304384  
GENERAL INFORMATION:



```
;; APPLICANT: Miller, Timothy J.
;; APPLICANT: Klepfer, Sharon
;; APPLICANT: Reed, Albert Paul
;; APPLICANT: Jones, Elaine V.
;; TITLE OF INVENTION: Compositions and Methods for Vaccination
;; AGAINST CORONAVIRUSES
;; NUMBER OF SEQUENCES: 48
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SmithKline Beecham Corporation - Corporate
;; ADDRESSEE: Patents
;; STREET: 709 Swedeland Road
;; CITY: King of Prussia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19406-2799
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/04384
;; FILING DATE: 19930507
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/882,171
;; FILING DATE: 08-MAY-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/698,927
;; FILING DATE: 13-MAY-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/613,066
;; FILING DATE: 14-NOV-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Schreck, Patricia A.
;; REGISTRATION NUMBER: 33,777
;; REFERENCE/DOCKET NUMBER: SBC H85009-1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 270-5015
;; TELEFAX: (215) 270-5090
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1454 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US93-04384-2

Query Match 9.58; Score 71.5; DB 5; Length 1454;
Best Local Similarity 27.08; Pred. No. 17;
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

Qy 36 LTRLYDKATLIWNGNAVSGLDLNNFFDTLPSSSEFOVNMILD-----CQPVHEQATQS 88
Db 211 VTLVYRSRSTATWEYSAAAYQGVSNF-----TYIKLNTNGLKTYELCED-YEHCTGY 263

Qy 89 QTTVLVWTSQGVKEDGNKQHFNFNLLTAOST-----PN-----NTVWKIAS 131
Db 264 ATNVFAPTSGGYIPDG---FSFNNWFLTNSSTFVSGRFTVNOPLLINCLWPVS 315

RESULT 15
PCT-US93-04384-8
; Sequence 8, Application PC/TUS9304384
; GENERAL INFORMATION:
; APPLICANT: Miller, Timothy J.
; APPLICANT: Klepfer, Sharon
; APPLICANT: Reed, Albert Paul
; APPLICANT: Jones, Elaine V.
; TITLE OF INVENTION: Compositions and Methods for Vaccination
; AGAINST CORONAVIRUSES
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
```

```
;; ADDRESSEE: SmithKline Beecham Corporation - Corporate
;; ADDRESSEE: Patents
;; STREET: 709 Swedeland Road
;; CITY: King of Prussia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19406-2799
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/04384
;; FILING DATE: 19930507
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/882,171
;; FILING DATE: 08-MAY-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/698,927
;; FILING DATE: 13-MAY-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/613,066
;; FILING DATE: 14-NOV-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Schreck, Patricia A.
;; REGISTRATION NUMBER: 33,777
;; REFERENCE/DOCKET NUMBER: SBC H85009-1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 270-5015
;; TELEFAX: (215) 270-5090
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1454 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US93-04384-8

Query Match 9.58; Score 71.5; DB 5; Length 1454;
Best Local Similarity 27.08; Pred. No. 17;
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

Qy 36 LTRLYDKATLIWNGNAVSGLDLNNFFDTLPSSSEFOVNMILD-----CQPVHEQATQS 88
Db 211 VTLVYRSRSTATWEYSAAAYQGVSNF-----TYIKLNTNGLKTYELCED-YEHCTGY 263

Qy 89 QTTVLVWTSQGVKEDGNKQHFNFNLLTAOST-----PN-----NTVWKIAS 131
Db 264 ATNVFAPTSGGYIPDG---FSFNNWFLTNSSTFVSGRFTVNOPLLINCLWPVS 315

Search completed: March 4, 2003, 15:06:27
Job time : 7.92282 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:58:57 ; Search time 3.39909 Seconds  
(without alignments)  
1761.695 Million cell updates/sec

Title: US-09-763-902B-7  
Perfect score: 749  
Sequence: 1 MATSLDKFTYDQACRAAE.....NTWKIASDCFRFDWSSS 142

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 727   | 97.1        | 162    | 10 | US-09-764-864-1180  |
| 2          | 584   | 78.0        | 164    | 10 | US-09-764-864-1181  |
| 3          | 83.5  | 11.1        | 1127   | 10 | US-09-815-242-5373  |
| 4          | 83.5  | 11.1        | 1158   | 10 | US-09-815-242-12522 |
| 5          | 78.5  | 10.5        | 450    | 9  | US-10-156-238-11    |
| 6          | 78.5  | 10.5        | 450    | 10 | US-09-795-693-11    |
| 7          | 72    | 9.6         | 576    | 10 | US-09-925-297-787   |
| 8          | 71.5  | 9.5         | 748    | 10 | US-09-854-799-24    |
| 9          | 71.5  | 9.5         | 748    | 10 | US-09-854-799-28    |
| 10         | 71.5  | 9.5         | 748    | 10 | US-09-854-799-34    |
| 11         | 71.5  | 9.5         | 1454   | 10 | US-09-854-799-22    |
| 12         | 71.5  | 9.5         | 1454   | 10 | US-09-854-799-26    |
| 13         | 68.5  | 9.1         | 477    | 10 | US-09-925-301-977   |
| 14         | 67.5  | 9.0         | 1179   | 10 | US-09-815-242-13608 |
| 15         | 66.5  | 8.9         | 498    | 10 | US-09-946-239-8     |
| 16         | 66    | 8.8         | 266    | 9  | US-09-738-626-6904  |
| 17         | 66    | 8.8         | 323    | 10 | US-09-943-002-12    |
| 18         | 65.5  | 8.7         | 5795   | 10 | US-09-815-242-12610 |
| 19         | 65    | 8.7         | 274    | 10 | US-09-815-242-5621  |

|    |      |     |      |    |                     |                    |
|----|------|-----|------|----|---------------------|--------------------|
| 20 | 65   | 8.7 | 282  | 10 | US-09-815-242-12241 | Sequence 12241, A  |
| 21 | 65   | 8.7 | 341  | 10 | US-09-877-804-4     | Sequence 4, Appl1  |
| 22 | 65   | 8.7 | 674  | 10 | US-09-877-804-3     | Sequence 3, Appl1  |
| 23 | 65   | 8.7 | 700  | 10 | US-09-877-804-2     | Sequence 2, Appl1  |
| 24 | 65   | 8.7 | 1454 | 10 | US-09-854-799-30    | Sequence 32, Appl1 |
| 25 | 64.5 | 8.6 | 302  | 9  | US-09-885-303A-20   | Sequence 20, Appl1 |
| 26 | 64   | 8.5 | 883  | 10 | US-09-815-242-13684 | Sequence 13684, A  |
| 27 | 63.5 | 8.5 | 475  | 9  | US-10-076-137-4     | Sequence 4, Appl1  |
| 28 | 63   | 8.4 | 351  | 9  | US-09-829-378-5     | Sequence 5, Appl1  |
| 29 | 63   | 8.4 | 351  | 9  | US-10-045-792-9     | Sequence 9, Appl1  |
| 30 | 62.5 | 8.3 | 1805 | 9  | US-09-820-843A-73   | Sequence 73, Appl1 |
| 31 | 62   | 8.3 | 612  | 8  | US-08-910-386A-2    | Sequence 2, Appl1  |
| 32 | 62   | 8.3 | 1025 | 8  | US-08-910-386A-7    | Sequence 7, Appl1  |
| 33 | 62   | 8.3 | 1289 | 9  | US-09-932-145-11    | Sequence 11, Appl1 |
| 34 | 61.5 | 8.2 | 311  | 10 | US-09-764-864-1543  | Sequence 1543, Ap  |
| 35 | 61.5 | 8.2 | 2835 | 10 | US-09-885-535-4     | Sequence 4, Appl1  |
| 36 | 61   | 8.1 | 360  | 10 | US-09-801-368-156   | Sequence 156, App  |
| 37 | 61   | 8.1 | 883  | 10 | US-09-815-242-13382 | Sequence 13382, A  |
| 38 | 61   | 8.1 | 2471 | 12 | US-10-116-048-4     | Sequence 4, Appl1  |
| 39 | 60.5 | 8.1 | 731  | 9  | US-09-881-579-2     | Sequence 2, Appl1  |
| 40 | 60.5 | 8.1 | 894  | 12 | US-10-060-332-4     | Sequence 4, Appl1  |
| 41 | 60   | 8.0 | 592  | 10 | US-09-765-272-222   | Sequence 222, App  |
| 42 | 59.5 | 7.9 | 247  | 9  | US-09-738-626-5773  | Sequence 5773, Ap  |
| 43 | 59.5 | 7.9 | 260  | 9  | US-09-738-626-4364  | Sequence 4364, Ap  |
| 44 | 59.5 | 7.9 | 394  | 9  | US-09-738-626-4331  | Sequence 4331, Ap  |
| 45 | 59.5 | 7.9 | 402  | 10 | US-09-888-243-14    | Sequence 14, Appl  |

## ALIGNMENTS

RESULT 1  
US-09-764-864-1180  
; Sequence 1180, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1180  
; LENGTH: 162  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-864-1180

|                       |                 |                                                          |               |             |
|-----------------------|-----------------|----------------------------------------------------------|---------------|-------------|
| Query Match           | 97.1%           | Score 727;                                               | DB 10;        | Length 162; |
| Best Local Similarity | 100.0%          | Pred. No. 6e-75;                                         | Mismatches 0; | Indels 0;   |
| Matches 137;          | Conservative 0; | 0;                                                       | Gaps 0;       |             |
| QY                    | 6               | DFKTYDQACRAAEFEVNIYETMDKRRRALTRYLDKATLIWNGNAVSGLDALNFFDT | 65            |             |
| Db                    | 26              | DFKTYDQACRAAEFEVNIYETMDKRRRALTRYLDKATLIWNGNAVSGLDALNFFDT | 85            |             |
| QY                    | 66              | LPSEFQVNMDCQPVHEQATQSQTTLVVTSGTVFDGKQHFNFQNLTAQSTPNNT    | 125           |             |
| Db                    | 86              | LPSEFQVNMDCQPVHEQATQSQTTLVVTSGTVFDGKQHFNFQNLTAQSTPNNT    | 145           |             |
| QY                    | 126             | VWKIASDCFRFDWSSS 142                                     |               |             |
| Db                    | 146             | VWKIASDCFRFDWSSS 162                                     |               |             |

RESULT 2  
US-09-764-864-1181  
; Sequence 1181, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

PRIORITY FILING DATE: 2008-08-20

; PRIOR FILING DATE: 2000-12-22.

PRIOR FILING DATE: 2000-12-22.

10

100

100

Db 129 DFKYA-VTPEHLESTFI-----ADKNMQRTRKKNLEADIALIYREFEQRIONEFITGEDA 181  
QY 59 LNNFFDTLPSSFEQVNMDCQPVHEQATQSTTV-LVVTSGTVKF-----DGNKQ 107  
Db 182 LQYFIDCHPKSEW-----LKCADIYDGFHNFSTIEYLIKGLIKYAKSVTIIITLTDGNHD 237  
QY 108 HF 109  
Db 238 QF 239

RESULT 5  
US-10-156-239-11  
; Sequence 11, Application US/10156239  
; Publication No. US20030036074A1  
; GENERAL INFORMATION:  
; APPLICANT: Glucksmann, Maria A.  
; TITLE OF INVENTION: No. US20030036074A1el Nucleic Acid Sequences Encoding Human Trans  
; TITLE OF INVENTION: Ubiquitin Conjugating Enzyme-Like Molecule, A Hu  
; TITLE OF INVENTION: Ubiquitin Conjugating Enzyme-Like Molecule, and Uses Therefor  
; FILE REFERENCE: 35800/247645  
; CURRENT APPLICATION NUMBER: US/10/156,239  
; CURRENT FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: 09/795,693  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/185,906  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: 09/809,557  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/192,018  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: 09/808,568  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 60/191,790  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: 09/808,767  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/191,781  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 11  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-156-239-11

Query Match 10.5%; Score 78.5; DB 9; Length 450;  
Best Local Similarity 29.8%; Pred. No. 0.65;  
Matches 31; Conservative 14; Mismatches 30; Indels 29; Gaps 6;  
QY 19 EEFVNIYYET-----MDKRRRALTRLYL-----DKATLIWNGNAVSGLDALNNFFDTL 65  
Db 52 DQTLNVTLEVTLSQIIDMDERNQVLT-LYLWIRQEWTDAYLRWDPNAYGGGLDAIR----- 105  
QY 66 LPSS-----EFQVNMDCQPVHEQATQSTTV-LVVTSGTVKF 103  
Db 106 IPSSLVWRPDIYLNKADAQP-----PGSASTNVVLRHDGAVRD 145

RESULT 6  
US-09-795-693-11  
; Sequence 11, Application US/09795693  
; Patent No. US20020068710A1  
; GENERAL INFORMATION:  
; APPLICANT: Glucksmann, Maria A.  
; TITLE OF INVENTION: 20685, 579, 17114, 23821, 33894, and  
; TITLE OF INVENTION: 32613, No. US20020068710A1el Human Transporters  
; FILE REFERENCE: 35800/209292  
; CURRENT APPLICATION NUMBER: US/09/795,693  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/185,906

Db 129 DFKYA-VTPEHLESTFI-----ADKNMQRTRKKNLEADIALIYREFEQRIONEFITGEDA 181  
QY 59 LNNFFDTLPSSFEQVNMDCQPVHEQATQSTTV-LVVTSGTVKF-----DGNKQ 107  
Db 182 LQYFIDCHPKSEW-----LKCADIYDGFHNFSTIEYLIKGLIKYAKSVTIIITLTDGNHD 237  
QY 108 HF 109  
Db 238 QF 239

RESULT 7  
US-09-925-297-787  
; Sequence 787, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: P105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 787  
; LENGTH: 576  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-297-787

Query Match 9.6%; Score 72; DB 10; Length 576;  
Best Local Similarity 21.2%; Pred. No. 4.9;  
Matches 31; Conservative 22; Mismatches 37; Indels 56; Gaps 6;  
QY 7 FKYIVDQACRAAEFVNIYYETMDKRRRALTRLYLDKATLIWNGNAVSGLDALNNFFDTL 65  
Db 53 FSDATSKAHOLEEDIVSVTHKAIEKE-----TELDSLK---DKL 88  
QY 67 PSSEFQVNMDCQ-----PVHEQATQSTTV-LVVTSGTVK-FDGNKQ---HF- 109  
Db 89 KKAQHEREQLECOLKTEKDEKELYKVHLKNTENKLNSEVOTLKNLDGKESVITFK 148  
QY 110 -----FNQNFLLTAQS 120  
Db 149 EEIGRLQLCLAEKENLQRTFLTTSS 174

RESULT 8  
US-09-854-799-24  
; Sequence 24, Application US/09854799  
; Patent No. US20020115064A1  
; GENERAL INFORMATION:  
; APPLICANT: SmithKline Beecham, Corporation  
; TITLE OF INVENTION: Recombinant Feline Coronavirus S  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road

Wed Mar 5 08:37:56 2003

CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/854,799  
FILING DATE: 14-May-2001  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/392,459  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-09-854-799-24  
Query Match 9.5%; Score 71.5; DB 10; Length 748;  
Best Local Similarity 27.0%; Pred. No. 7.8;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;  
QY 36 LTRLXLDKATLWNGAVSGLDALNNFFDTLPSSFEQVNMLD-----CQPVHEQATQS 88  
DB 211 VTLVSRSTATWEYSAAAYQGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263  
QY 89 QTVLVVTSVTGKFGKQKHFQNFLLTAQST-----PN-----NTVWKIAS 131  
DB 264 ATNVFAPTSGGYIPDG---FSFNNWFLNLSSTFVSGRFTVNTQPLLNCMLPVPVS 315  
RESULT 9  
US-09-854-799-28  
Sequence 28, Application US/09854799  
Patent No. US20020115064A1  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/854,799  
FILING DATE: 14-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/392,459  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 28:  
US-09-854-799-28  
Query Match 9.5%; Score 71.5; DB 10; Length 748;  
Best Local Similarity 27.0%; Pred. No. 7.8;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;  
QY 36 LTRLXLDKATLWNGAVSGLDALNNFFDTLPSSFEQVNMLD-----CQPVHEQATQS 88  
DB 211 VTLVSRSTATWEYSAAAYQGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263  
QY 89 QTVLVVTSVTGKFGKQKHFQNFLLTAQST-----PN-----NTVWKIAS 131  
DB 264 ATNVFAPTSGGYIPDG---FSFNNWFLNLSSTFVSGRFTVNTQPLLNCMLPVPVS 315  
RESULT 10  
US-09-854-799-34  
Sequence 34, Application US/09854799  
Patent No. US20020115064A1  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/854,799  
FILING DATE: 14-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/392,459  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids

TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-09-854-799-34:

Query Match 9.5%; Score 71.5; DB 10; Length 748;  
Best Local Similarity 27.0%; Pred. No. 7.8;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

QY 36 LTRLYDKATLIWNGNAVGDLNNEFDTLPSSEFQVNMID-----CPVHEQATQS 88  
Db 211 VTLIYRSRSTATWEYSAAAYQGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263

QY 89 OTTVLVVTSGTVKFDGNKQHFNFQNLTAOST-----PN-----NTVVKIAS 131  
Db 264 ATNVFAPTSGGYIPDG---FSFNWFLTNSSTFVSGRFTVNPQLINCLMPVPS 315

## RESULT 11

US-09-854-799-22  
Sequence 22, Application US/09854799  
Patent No. US20020115064A1

GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
Proteins

NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA

COUNTRY: USA

ZIP: 19406-2799

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/854,799

FILING DATE: 14-May-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/392,459

FILING DATE: <Unknown>

APPLICATION NUMBER: US 07/613,066

FILING DATE: 14-NOV-1990

ATTORNEY/AGENT INFORMATION:

NAME: King, William T.

REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: SEC 14532B

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 270-5015

TELEFAX: (215) 270-5090

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 1454 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 22:

## US-09-854-799-22

Query Match 9.5%; Score 71.5; DB 10; Length 1454;  
Best Local Similarity 27.0%; Pred. No. 19;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

QY 36 LTRLYDKATLIWNGNAVGDLNNEFDTLPSSEFQVNMID-----CPVHEQATQS 88  
Db 211 VTLIYRSRSTATWEYSAAAYQGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263

QY 89 OTTVLVVTSGTVKFDGNKQHFNFQNLTAOST-----PN-----NTVVKIAS 131  
Db 264 ATNVFAPTSGGYIPDG---FSFNWFLTNSSTFVSGRFTVNPQLINCLMPVPS 315

## RESULT 12

US-09-854-799-26

Sequence 26, Application US/09854799

Patent No. US20020115064A1

GENERAL INFORMATION:

APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
Proteins

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESSES:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-2799

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/854,799

FILING DATE: 14-May-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/392,459

FILING DATE: <Unknown>

APPLICATION NUMBER: US 07/613,066

FILING DATE: 14-NOV-1990

ATTORNEY/AGENT INFORMATION:

NAME: King, William T.

REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: SEC 14532B

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 270-5015

TELEFAX: (215) 270-5090

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 1454 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Query Match 9.5%; Score 71.5; DB 10; Length 1454;  
Best Local Similarity 27.0%; Pred. No. 19;  
Matches 31; Conservative 14; Mismatches 41; Indels 29; Gaps 6;

QY 36 LTRLYDKATLIWNGNAVGDLNNEFDTLPSSEFQVNMID-----CPVHEQATQS 88  
Db 211 VTLIYRSRSTATWEYSAAAYQGVSNF-----TYKLNNTNGLKTYELCED-YEHCTGY 263

QY 89 OTTVLVVTSGTVKFDGNKQHFNFQNLTAOST-----PN-----NTVVKIAS 131  
Db 264 ATNVFAPTSGGYIPDG---FSFNWFLTNSSTFVSGRFTVNPQLINCLMPVPS 315

## RESULT 13

US-09-925-301-977

Sequence 977, Application US/09925301

Patent No. US20020052308A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA106

CURRENT APPLICATION NUMBER: US/09/925,301

CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 977  
LENGTH: 477  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (471)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (473)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-977

Query Match 9.1%; Score 68.5; DB 10; Length 477;  
Best Local Similarity 22.9%; Pred. No. 9.5;  
Matches 30; Conservative 22; Mismatches 58; Indels 21; Gaps 5;  
QY 20 EFVNIYETMDKRRRLTRLYLDKATLIWNG-----NAVSGLDALNFFDTLPSSSEF 71  
DB 25 EFVQYITLQAPDMLHRYGYNSSVHGGLDSNGKPADAVYQKKEIHR-----KVM 77  
QY 72 QVNNLDC-----QPVEHQATQSOTVLVVTSGTVKFDGNKQHFNFQNLTAQSTPNNTW 127  
DB 78 SQNFTHCKIRHVDAHATLND-GVVVGVNGLLSNNQALRRRMTQFVLAPEGSVANKFY 136  
QY 128 KIASDCFRFOD 138  
DB 137 -VHNDIFRYOD 146

RESULT 14  
US-09-815-242-13608  
Sequence 13608, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 13608  
LENGTH: 1179

TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13608  
Query Match 9.0%; Score 67.5; DB 10; Length 1179;  
Best Local Similarity 23.7%; Pred. No. 40;  
Matches 37; Conservative 19; Mismatches 73; Indels 27; Gaps 7;  
QY 1 MATSLDFKTYVDQACRAAEFVNIYETMDKRRRLTRLYLDKATLIWNGNAVSGLDALN 60  
DB 621 LATTAIFDT-VEHAREARKQ---VRIOV-----RMVT---LOGTELTGSGYAGGANRQN 668  
QY 61 NFFDTLPSSSE-----FOVNMILDCQPVHEQATQSOTVLVVTSGTVKFDGNKQHFNFQ 113  
DB 669 NSIFIKPELEQLQKETAEEASLDSSEVALKTLQDEMARLTESLEAIKSGEQARIQEQ 728  
QY 114 FLTAQSTPN-----NTVWKIAS---DCFRFQDWSS 141  
DB 729 LFLAYQOTSQOVVELETMLKQEEIDRLSEGDWQA 764

RESULT 15  
US-09-946-239-8  
Sequence 8, Application US/09946239  
Patent No. US20020044945A1  
GENERAL INFORMATION:  
APPLICANT: Barr, Margaret C.  
TITLE OF INVENTION: No. US20020044945A1 Feline Immunodeficiency Virus Nucleotide  
TITLE OF INVENTION: Polypeptide Sequences  
FILE REFERENCE: 18617.0059  
CURRENT APPLICATION NUMBER: US/09/946,239  
CURRENT FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: US 09/238,303, US 60/072,927  
PRIOR FILING DATE: 1999-01-28, 1998-01-29  
NUMBER OF SEQ ID NOS: 17  
SEQ ID NO 8  
LENGTH: 498  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: protein encoded by the gag gene of a recombinant viral  
OTHER INFORMATION: clone constructed from the genomic DNA of a Pallas's cat felin  
OTHER INFORMATION: immunodeficiency virus  
US-09-946-239-8

Query Match 8.9%; Score 66.5; DB 10; Length 498;  
Best Local Similarity 21.0%; Pred. No. 17;  
Matches 33; Conservative 27; Mismatches 56; Indels 41; Gaps 6;  
QY 6 DEKTYVDQACRAAEFVNIYETMDKRRRLTRLYLDKATLIWNG-----AVSGLDALN 60  
DB 307 DYASFIDR-----LYAQIDQONSPEVKIYLLKQSLSLANANPECKKAMSHLPES 356  
QY 61 NFFDTL-----PSSEFQVNMILDCQPVHEQATQSOTVLVVTSGTVKFDGNKQH----- 108  
DB 357 TLEEKLRACQEVGSGTSYKMNML-----AQAQOOSQVQVQOGRGKPGQNNRRPGOSLK 410  
QY 109 FFN-----QNFLTAQSTPNNTVWKIASDCFRFQ 137  
DB 411 CFNCGRPGHLARNCRAPRKCKGKAGHIATDCWDMQ 447

Search completed: March 4, 2003, 15:07:28  
Job time : 5.39909 secs

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:50:41 ; Search time 21.187 seconds  
(without alignments)  
1559.740 Million cell updates/sec

Title: US-09-763-902b-8

Perfect score: 1280

Sequence: 1 MSVSLPLTWVWRDWIGH.....LRQLRQPSQAQVAPQCC 248

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

|     |                                                  |
|-----|--------------------------------------------------|
| 1:  | /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:* |
| 2:  | /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:* |
| 3:  | /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:* |
| 4:  | /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:* |
| 5:  | /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:* |
| 6:  | /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:* |
| 7:  | /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:* |
| 8:  | /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:* |
| 9:  | /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:* |
| 10: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:* |
| 11: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:* |
| 12: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:* |
| 13: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:* |
| 14: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:* |
| 15: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:* |
| 16: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:* |
| 17: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:* |
| 18: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:* |
| 19: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:* |
| 20: | /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:* |
| 21: | /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:* |
| 22: | /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:* |
| 23: | /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 1280  | 100.0       | 248    | 21 | Human protein tran |
| 2          | 1147  | 89.6        | 222    | 22 | Novel signal trans |
| 3          | 1147  | 89.6        | 225    | 22 | Novel signal trans |
| 4          | 1070  | 83.6        | 213    | 21 | dRab4 amino acid s |
| 5          | 1065  | 83.2        | 213    | 21 | rRab4b amino acid  |
| 6          | 1064  | 83.1        | 213    | 21 | Human Rab4b protei |
| 7          | 828   | 64.7        | 213    | 22 | Drosophila melanog |
| 8          | 710   | 55.5        | 143    | 21 | Amino acid sequenc |
| 9          | 645.5 | 50.4        | 215    | 21 | Amino acid sequenc |
| 10         | 645.5 | 50.4        | 215    | 21 | Human ORFX ORF2487 |

|    |       |      |     |    |                    |
|----|-------|------|-----|----|--------------------|
| 11 | 645.5 | 50.4 | 215 | 22 | Human protein sequ |
| 12 | 645.5 | 50.4 | 307 | 22 | Novel signal trans |
| 13 | 645.5 | 50.4 | 312 | 21 | Lung cancer associ |
| 14 | 622.5 | 48.6 | 215 | 22 | Drosophila melanog |
| 15 | 615.5 | 48.1 | 174 | 22 | Novel human diagno |
| 16 | 550   | 43.0 | 211 | 21 | Arabidopsis thalia |
| 17 | 550   | 43.0 | 256 | 21 | Arabidopsis thalia |
| 18 | 546   | 42.7 | 216 | 21 | Amino acid sequenc |
| 19 | 546   | 42.7 | 216 | 22 | Human polypeptide  |
| 20 | 546   | 42.7 | 236 | 22 | Human polypeptide  |
| 21 | 545.5 | 42.6 | 213 | 22 | Drosophila melanog |
| 22 | 543   | 42.4 | 212 | 22 | Human wound healin |
| 23 | 539   | 42.1 | 212 | 11 | Protein with cell  |
| 24 | 539   | 42.1 | 212 | 21 | Amino acid sequenc |
| 25 | 539   | 42.1 | 212 | 22 | Mouse wound healin |
| 26 | 539   | 42.1 | 212 | 23 | Rat potassium chan |
| 27 | 539   | 42.1 | 212 | 23 | Mouse ischaemic co |
| 28 | 535   | 41.8 | 210 | 21 | Zea mays protein f |
| 29 | 534   | 41.7 | 211 | 21 | Arabidopsis thalia |
| 30 | 534   | 41.7 | 211 | 21 | Arabidopsis thalia |
| 31 | 530   | 41.4 | 209 | 22 | Human G protein    |
| 32 | 511.5 | 40.0 | 264 | 21 | Arabidopsis thalia |
| 33 | 503   | 39.3 | 205 | 21 | Arabidopsis thalia |
| 34 | 462   | 36.1 | 148 | 22 | Novel human diagno |
| 35 | 453.5 | 35.4 | 234 | 21 | Arabidopsis thalia |
| 36 | 452.5 | 35.4 | 214 | 22 | Drosophila melanog |
| 37 | 452   | 35.3 | 207 | 21 | Arabidopsis thalia |
| 38 | 450   | 35.2 | 224 | 21 | Arabidopsis thalia |
| 39 | 449   | 35.1 | 224 | 21 | Arabidopsis thalia |
| 40 | 442.5 | 34.6 | 233 | 21 | Arabidopsis thalia |
| 41 | 442   | 34.5 | 194 | 21 | Human secreted pro |
| 42 | 439   | 34.3 | 217 | 21 | Arabidopsis thalia |
| 43 | 439   | 34.3 | 217 | 21 | Arabidopsis thalia |
| 44 | 438.5 | 34.3 | 225 | 21 | Arabidopsis thalia |
| 45 | 437.5 | 34.2 | 233 | 21 | Pea light-repressi |

#### ALIGNMENTS

#### RESULT 1

AA1982324

ID AA1982324 standard; Protein; 248 AA.

XX AA1982324;

XX DT 19-JUN-2000 (first entry)

XX DE Human protein transport molecule (PTAM) SEQ ID NO:8.

XX KW Human; protein transport molecule; PTAM; diagnosis; cytostatic;

XX KW antiarthritic; antidiabetic; antidiabetic; immunosuppressant; antiarteriosclerotic;

XX KW dermatologic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;

XX KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; osteoporotic;



XX WPI; 2000-256642/22.  
 DR N-PSDB; AAA08042.  
 XX  
 PT New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis  
 XX  
 PS Claim 1; Page 68-69; 75pp; English.  
 XX  
 CC AAA08035 to AAA08042 encode the human protein transport-associated  
 CC molecules (PTAMs) given in AY82317 to AY82324. The PTAMs have  
 CC cytosolic, antihistaminic, antidiabetic, immunosuppressant,  
 CC antihistaminic, antiallergic, antidiabetic, antilipemic,  
 CC antirheumatic, osteopathic, dermatological, antianemic, antipsoriatic,  
 CC hepatotropic, antitumor, antineoplastic and antihIV activities, and  
 CC regulate protein transport. PTAM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PTAM. PTAM polynucleotides are useful for  
 CC diagnosing conditions associated with PTAM, comprising detecting PTAM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the  
 CC biological sample. Diseases prevented, treated or diagnosed include cell  
 CC proliferative disorders such as cancers, immune disorders, secretory  
 CC disorders and other conditions associated with abnormal vesicle  
 CC trafficking, such as allergies, asthma, urticaria and autoimmune  
 CC haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as  
 CC a targeting or delivery mechanism for bringing pharmaceutical agents  
 CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related  
 CC disorders. PTAM, its catalytic or immunogenic fragments are useful for  
 CC drug screening using libraries of compounds. PTAM polynucleotides are  
 CC useful for generating hybridisation probes useful in mapping the  
 CC naturally occurring genomic sequences.  
 XX  
 SQ Sequence 248 AA;

Query Match 100.0%; Score 1280; DB 21; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-133;  
 Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSVSLPLTVVRRDWTGIIHLSYLSLPVIGIPDFGSIWSDFLFKFLVIGSAGTKSCLL 60  
 DB 1 MSVSLPLTVVRRDWTGIIHLSYLSLPVIGIPDFGSIWSDFLFKFLVIGSAGTKSCLL 60  
 OY 61 HQTENFKQDSNHTIGVEFGSRVNVGKTVKQIWDTAGOERFSVTRSYRGAAGAL 120  
 DB 61 HQTENFKQDSNHTIGVEFGSRVNVGKTVKQIWDTAGOERFSVTRSYRGAAGAL 120  
 OY 121 LVYDITSRETYNSLAAMLTARTLASPNVILCGNKDLPDREVTFLFASRAQENEL 180  
 DB 121 LVYDITSRETYNSLAAMLTARTLASPNVILCGNKDLPDREVTFLFASRAQENEL 180  
 OY 181 MFLTSALTGENVEAFKLCARTILNKIDSGELDPERMGSGIQYGDASLRQLRQPSAQA 240  
 DB 181 MFLTSALTGENVEAFKLCARTILNKIDSGELDPERMGSGIQYGDASLRQLRQPSAQA 240  
 OY 241 VAPQPCGC 248  
 DB 241 VAPQPCGC 248

RESULT 2  
 AAU17547  
 ID AAU17547 standard; Protein; 222 AA.  
 XX  
 AC AAU17547;  
 XX  
 XX 07-NOV-2001 (first entry)  
 DT  
 XX Novel signal transduction pathway protein, Seq ID 1112.  
 DE  
 XX Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
 KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
 KW immune system disorder; rheumatoid arthritis; inflammatory condition;

KW organ transplant rejection; infection; hepatitis C; blood disorder;  
 KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
 KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
 KW reproductive system; gastrointestinal; liver disorder; AIDS;  
 XX acquired immune deficiency syndrome.  
 OS Homo sapiens.  
 XX  
 XX WO200154733-A1.  
 XX 02-AUG-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-US01312.  
 XX  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 14-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.

```
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465460/50.
XX N-PSDB; AAS27464.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
XX Claim 1; SEQ ID No 1112; 880pp; English.
XX
XX The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Query Match 89.6%; Score 1147; DB 22; Length 222;
Best Local Similarity 100.0%; Pred. No. 3.6e-118;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 SLPVGIPDFGSIWSDFLFKFLVIGSAGTGKSKLLHQFTENKFKODSNHTIGVEGSRVYN 86
Db 1 SLPVGIPDFGSIWSDFLFKFLVIGSAGTGKSKLLHQFTENKFKODSNHTIGVEGSRVYN 60
QY 87 VGGKTVKLQIWDTAGQERFSRVSRYRGAAGALLVYDITSRETYNSLAAMLTARTLAS 146
Db 61 VGGKTVKLQIWDTAGQERFSRVSRYRGAAGALLVYDITSRETYNSLAAMLTARTLAS 120
QY 147 PNIVVILCGNKKDLPDREVTFLFASRFAQENELMFLFLETSGENVEEAFKLCARTILN 206
Db 121 PNIVVILCGNKKDLPDREVTFLFASRFAQENELMFLFLETSGENVEEAFKLCARTILN 180
QY 207 KIDSGELDPERMGSGIOYGDASLRQLRQPRSAQAVAPQPCG 248
Db 181 KIDSGELDPERMGSGIOYGDASLRQLRQPRSAQAVAPQPCG 222

RESULT 3
AAU17127
ID AAU17127 standard; Protein; 225 AA.
XX
XX AAU17127;
XX
XX 07-NOV-2001 (first entry)
XX
```

DE Novel signal transduction pathway protein, Seq ID 692.  
XX Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
KW antinflammatory; anti-HIV; antibacterial; antinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS;  
XX acquired immune deficiency syndrome.  
OS Homo sapiens.  
XX  
XX WO200154733-A1.  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01312.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 04-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-465460/50.  
 XX N-PSDB; AAS27044.  
 DR  
 DR Novel polypeptides useful for diagnosing, treating, preventing and/or  
 PT prognosing disorders related to the proteins, including cancers, immune  
 PT disorders and neuronal disorders  
 XX  
 XX Claim 1; SEQ ID No 692; 880pp; English.  
 PS  
 XX The invention relates to novel isolated polypeptides (I), and  
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
 CC diagnosing, preventing and treating diseases including immune system  
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
 CC transplant rejections and graft versus host disease, infectious diseases  
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
 CC disorders, primary haematopoietic disorders, hyperproliferative  
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal  
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal  
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders,  
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
 CC Addison's disease), reproductive system disorders, gastrointestinal  
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
 CC as stimulators of B-cell responsiveness to pathogens, activators of  
 CC T-cells, to induce higher affinity antibodies, and as a means to induce  
 CC tumour proliferation in pathologies e.g. acquired immune deficiency  
 CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction  
 CC pathway protein, amino acid sequences of the invention.  
 XX

Query Match 89.6%; Score 1147; DB 22; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-118;  
 Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 SLPVGPIDFGSIWDSDFELFKFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEGSRVWN 86  
 Db 4 SLPVGPIDFGSIWDSDFELFKFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEGSRVWN 63  
 QY 87 VGGKTVKLIQIWDTAGQERFRSRTSYRGAAGALLVYDITSRETYNSLAAMLTDARTLAS 146  
 Db 64 VGGKTVKLIQIWDTAGQERFRSRTSYRGAAGALLVYDITSRETYNSLAAMLTDARTLAS 123  
 QY 147 PNIVILCGNKKOLDPEREVTFLASRFAQENELMFLFLETSAITGENVEEAFKLCARTILN 206  
 Db 124 PNIVILCGNKKOLDPEREVTFLASRFAQENELMFLFLETSAITGENVEEAFKLCARTILN 183  
 QY 207 KIDSGELDPERMGSQYGDASLRQLRQPSAQAVAPQCGC 248  
 Db 184 KIDSGELDPERMGSQYGDASLRQLRQPSAQAVAPQCGC 225

RESULT 4  
 AAB23762  
 ID AAB23762 standard; Protein; 213 AA.

XX AAB23762;  
 AC 12-JAN-2001 (first entry)  
 XX dRab4 amino acid sequence.  
 DE Human; Rab4b; Rab4.  
 XX Unidentified.  
 OS CN1257124-A.  
 PN 21-JUN-2000.  
 PD 11-DEC-1998; 98CN-0125320.  
 XX 11-DEC-1998; 98CN-0125320.  
 PR (UYFU-) UNIV FUDAN.  
 PA Yu L, Zhang H, Tu Q;  
 XX WPI; 2000-565932/53.  
 DR Novel human protein and coded sequence, preparation method and use -  
 XX Example 2; Fig 1; 22pp; Chinese.  
 PS The present invention describes a novel human gene nucleotide sequence  
 CC and protein designated Rab4b. The human Rab4b protein sequence shows  
 CC homology with the protein sequences drab4 and rRab4b (AAB23762 and  
 CC AAB23763, respectively). The present sequence represents the drab4 amino  
 CC acid sequence given in the present invention.  
 XX Sequence 213 AA;  
 SQ

Query Match 83.6%; Score 1070; DB 21; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-109;  
 Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 DFLFKFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEGSRVWNVGGKTVKLIQIWDTA 100  
 Db 6 DFLFKFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEGSRVWNVGGKTVKLIQIWDTA 65  
 QY 101 GQERFRSRTSYRGAAGALLVYDITSRETYNSLAAMLTDARTLASPNIVILCGNKKDL 160  
 Db 66 GQERFRSRTSYRGAAGALLVYDITSRETYNSLAAMLTDARTLASPNIVILCGNKKDL 125  
 QY 161 DPREVTFLEASRFAQENELMFLFLETSAITGENVEEAFKLCARTILNKIDSGELDPERMGS 220  
 Db 126 DPREVTFLEASRFAQENELMFLFLETSAITGENVEEAFKLCARTILNKIDSGELDPERMGS 185  
 QY 221 GQYGDASLRQLRQPSAQAVAPQCGC 248  
 Db 186 GQYGDASLRQLRQPSAQAVAPQCGC 213

RESULT 5  
 AAB23763  
 ID AAB23763 standard; Protein; 213 AA.  
 XX AAB23763;  
 XX 12-JAN-2001 (first entry)  
 DT rRab4b amino acid sequence.  
 DE Human; Rab4b; Rab4.  
 XX Unidentified.  
 OS CN1257124-A.  
 PN

XX PD 21-JUN-2000.  
 XX XX  
 XX 11-DEC-1998; 98CN-0125330.  
 XX XX  
 XX 11-DEC-1998; 98CN-0125330.  
 XX XX  
 XX (UYFU-) UNIV FUDAN.  
 XX PA  
 XX Yu L, Zhang H, Tu Q;  
 XX DR  
 XX WPI; 2000-565932/53.  
 XX XX  
 XX Novel human protein and coded sequence, preparation method and use.  
 XX PT  
 XX Example 2; Fig 1; 22pp; Chinese.  
 XX PS  
 XX The present invention describes a novel human gene nucleotide sequence  
 CC CC and protein designated Rab4b. The human Rab4b protein sequence shows  
 CC CC and protein designated Rab4b. The human Rab4b protein sequence shows  
 CC CC homology with the protein sequences dRab4 and rRab4b (AAB23762 and  
 CC CC AAB23763, respectively). The present sequence represents the rRab4b amino  
 CC CC acid-sequence given in the present invention.  
 XX XX  
 XX SQ Sequence 213 AA;  
 Query Match 83.28; Score 1065; DB 21; Length 213;  
 Best Local Similarity 99.58; Pred. No. 3.9e-109;  
 Matches 207; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 41 DFLKFLVIGSAGTCKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVVGGTKVKLQIWDTA 100  
 Db 6 DFLKFLVIGSAGTCKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVVGGTKVKLQIWDTA 65  
 QY 101 QERFRSVTRSYRGAAGALLVYDITSRETYNSLAALWLTARTLASPNIVVILCGNKKDL 160  
 Db 66 QERFRSVTRSYRGAAGALLVYDITSRETYNSLAALWLTARTLASPNIVVILCGNKKDL 125  
 QY 161 DPERVTFLEASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 220  
 Db 126 DPERVTFLEASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 185  
 QY 221 GIOYGDASLRQLRQPRSAQAQVAPQCGC 248  
 Db 186 GIOYGDASLRQLRQPRSAQAQVAPQCGC 213  
 RESULT 6  
 AAB23761.  
 ID AAB23761 standard; Protein; 213 AA.  
 AC AAB23761;  
 XX XX  
 XX 12-JAN-2001 (first entry)  
 XX XX  
 XX Human Rab4b protein sequence SEQ ID NO:4.  
 XX XX  
 XX Human; Rab4b; Rab4.  
 XX KW  
 XX Homo sapiens.  
 XX OS  
 XX CN1257124-A.  
 XX PN  
 XX 21-JUN-2000.  
 XX PD  
 XX 11-DEC-1998; 98CN-0125330.  
 XX PF  
 XX 11-DEC-1998; 98CN-0125330.  
 XX PR  
 XX (UYFU-) UNIV FUDAN.  
 XX PA  
 XX Yu L, Zhang H, Tu Q;  
 XX PI  
 XX WPI; 2000-565932/53.  
 XX CC

DR N-PSDB; AAA923348.  
 XX XX  
 XX Novel human protein and coded sequence, preparation method and use.  
 XX PT  
 XX Claim 4; Page 17; 22pp; Chinese.  
 XX PS  
 XX The present invention describes a novel human gene nucleotide sequence  
 CC CC and protein designated Rab4b. The human Rab4b protein sequence shows  
 CC CC and protein designated Rab4b. The human Rab4b protein sequence shows  
 CC CC homology with the protein sequences dRab4 and rRab4b (AAB23762 and  
 CC CC AAB23763, respectively). The present sequence represents human Rab4b.  
 XX XX  
 XX SQ Sequence 213 AA;  
 Query Match 83.18; Score 1064; DB 21; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 5e-109;  
 Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 42 FLKFLVIGSAGTCKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVVGGTKVKLQIWDTAG 101  
 Db 7 FLKFLVIGSAGTCKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVVGGTKVKLQIWDTAG 66  
 QY 102 QERFRSVTRSYRGAAGALLVYDITSRETYNSLAALWLTARTLASPNIVVILCGNKKDL 161  
 Db 67 QERFRSVTRSYRGAAGALLVYDITSRETYNSLAALWLTARTLASPNIVVILCGNKKDL 126  
 QY 162 PEREVTFLASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
 Db 127 PEREVTFLASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 186  
 QY 222 IOYGDASLRQLRQPRSAQAQVAPQCGC 248  
 Db 187 IOYGDASLRQLRQPRSAQAQVAPQCGC 213  
 RESULT 7  
 ABB61094  
 ID ABB61094 standard; Protein; 213 AA.  
 AC ABB61094;  
 XX XX  
 XX 26-MAR-2002 (first entry)  
 XX DT  
 XX Drosophila melanogaster polypeptide SEQ ID NO 10074.  
 XX DE  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 XX KW pharmaceutical.  
 XX OS  
 XX Drosophila melanogaster.  
 XX PN  
 XX WO200171042-A2.  
 XX PD  
 XX 27-SEP-2001.  
 XX XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX PF  
 XX 23-MAR-2000; 2000US-191637P.  
 XX PR  
 XX 11-JUL-2000; 2000US-0614150.  
 XX PA  
 XX (PEKE) PE CORP NY.  
 XX XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX PI  
 XX WPI; 2001-656860/75.  
 XX DR  
 XX N-PSDB; ABL05197.  
 XX XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT PT interactions.  
 XX XX  
 XX Disclosure; SEQ ID NO 10074; 21pp + Sequence Listing; English.  
 XX PS  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC CC capable of detecting 1000 or more genes from Drosophila. The invention is

XX The present sequence represents a mammalian RAB protein. RAB proteins  
CC constitute the largest family of small GTPases, with over 40 identified  
CC isoforms. RAB proteins contain four highly conserved peptide sequences  
CC involved in GTP binding and hydrolysis. Compositions comprising RAB  
CC, nucleic acid are useful for identifying homologous or related genes,

CC the gene in a biological specimen.

XX Sequence 215 AA;

Query Match 50.4%; Score 645.5; DB 21; Length 215;  
Best Local Similarity 58.0%; Pred. No. 8.9e-63;  
Matches 123; Conservative 33; Mismatches 45; Indels 11; Gaps 3;

QY 42 FLFKFLVIGSAGTGKSCLLHOFLENKFKDQSNHTIGVEGSRVNVGGTKVLIQIWDTAG 101  
DB 10 YIFKYIIIGDMGVGKSCLLHOFTEKFMADCPHTIGVEGTRIIIEVSGQIKLIQIWDTAG 69  
QY 102 QERFVTRSYRGAAGALVYDITSRSTYNSLAWLTDARTLASPNVIVLCGNKKDLD 161  
DB 70 QERFVTRSYRGAAGALVYDITSRSTYNSLAWLTDARTLASPNVIVLCGNKKDLD 129  
QY 162 PEREVTFLASRAQNEMLFLETSALTGENVEAFKLCARTILNKIDSGELDPERMGC 221  
DB 130 AQDVTYEEAKQFAENGGLFLASAKTGENVEDAFLEAKKIYQIQDGLDLNAESG 189  
QY 222 IQGDSASLRQLRPSQAQVA-PQP-----CGC 248  
DB 190 VQKPSA-----PQGRLTSEPPQREGCGC 215

RESULT 10

AAB42723  
ID AAB42723 standard; Protein; 215 AA.

XX AAB42723;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF2487 polypeptide sequence SEQ ID NO:4974.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-ESDS; RAC76932.

XX Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease

XX Claim 11; Page 4147; 5507pp; English.

XX AACT74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 215 AA;

Query Match 50.4%; Score 645.5; DB 21; Length 215;  
Best Local Similarity 58.0%; Pred. No. 8.9e-63;  
Matches 123; Conservative 33; Mismatches 45; Indels 11; Gaps 3;

QY 42 FLFKFLVIGSAGTGKSCLLHOFLENKFKDQSNHTIGVEGSRVNVGGTKVLIQIWDTAG 101  
DB 10 YIFKYIIIGDMGVGKSCLLHOFTEKFMADCPHTIGVEGTRIIIEVSGQIKLIQIWDTAG 69

QY 102 QERFVTRSYRGAAGALVYDITSRSTYNSLAWLTDARTLASPNVIVLCGNKKDLD 161  
DB 70 QERFVTRSYRGAAGALVYDITSRSTYNSLAWLTDARTLASPNVIVLCGNKKDLD 129

QY 162 PEREVTFLASRAQNEMLFLETSALTGENVEAFKLCARTILNKIDSGELDPERMGC 221  
DB 130 AQDVTYEEAKQFAENGGLFLASAKTGENVEDAFLEAKKIYQIQDGLDLNAESG 189

QY 222 IQGDSASLRQLRPSQAQVA-PQP-----CGC 248

DB 190 VQKPSA-----PQGRLTSEPPQREGCGC 215

RESULT 11

AAB94663  
ID AAB94663 standard; Protein; 215 AA.

XX AAB94663;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:15588.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 DR  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs  
 XX  
 PS Claim 8; SEQ ID 15588; 2537pp + CD.ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AA03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AA02446 to  
 CC AA095893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 215 AA;  
 Query Match 50.4%; Score 645.5; DB 22; Length 215;  
 Best Local Similarity 58.0%; Pred. No. 8.9e-63;  
 Matches 123; Conservative 33; Mismatches 45; Indels 11; Gaps 3;  
 QY 42 FLEKELVIGSAGTGRSCLLHQFIENKFKODSNHTIGVEGSRVNVGGTKVLQIWDTAG 101  
 DB 10 YIFKIIIGDMGVGSCLLHQFTKFKMADCPHTIGVEGTRIEVSGQKIKLQIWDTAG 69  
 QY 102 QERFRSVTSYRGAAGALLVYDITSRETYNSLAALWLTARTLASPNIVVILCGNKKDL 161  
 DB 70 QERFRAVTSYRGAAGALWVYDITRSTYNSLWLTARTLTNTPTNVTLLIGNKADLE 129  
 QY 162 PERVTFLEASRFAQENELMFLTSLTGTENVEAFKLCARTINKIDSGELDPERMGG 221  
 DB 130 AQDYTYEAKQFAENGLLFLEASAKTENVDEAFLEAKKIYQIQDGLDLNAESC 189  
 QY 222 IQYGDASLRQLRPSQAQAVA-PQP-----CGC 248  
 DB 190 VQHKPSA-----PQGRLTSEPOREGCGC 215  
 RESULT 12  
 AA017535  
 ID AA017535 standard; Protein; 307 AA.  
 XX  
 AC AA017535;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Novel signal transduction pathway protein, Seq ID 1100.  
 XX  
 KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
 KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;

KW Immune system disorder; rheumatoid arthritis; inflammatory condition;  
 KW organ transplant rejection; infection; hepatitis C; blood disorder;  
 KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
 KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
 KW reproductive system; gastrointestinal; liver disorder; AIDS;  
 KW acquired immune deficiency syndrome.  
 XX Homo sapiens.  
 XX WO200154733-A1.  
 XX PD 02-AUG-2001.  
 XX PF 17-JAN-2001; 2001WO-US01312.  
 XX PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225477.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.





XX DE Lung cancer associated polypeptide sequence SEQ ID 783.

XX DE Human; lung cancer associated protein; neuroprotective; cytostatic;

XX KW cardioactive; immunomodulatory; muscular active; vulnerary;

XX KW gastrointestinal; nephrotropic; antineoplastic; gynecological;

XX KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

XX KW proliferative disorder; wound healing; infectious disease.

XX OS Homo sapiens.

XX PN WO200005180-A2.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05918.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (ROSE/). ROSEN C A.

XX PI Ruben SM;

XX DR WPI: 2000-587514/55.

XX DR N-PSDB; AAF18321.

XX PT Lung cancer associated gene sequences, referred to as lung cancer

XX PT antigens, useful for treatment, prevention, and diagnosis of disorders

XX PT such as lung cancer.

XX PS Claim 11; Page 1315:1316; 1425pp; English.

XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer

XX CC associated proteins represented in AAF58106 - AAF58548. Lung cancer

XX CC associated proteins and polynucleotide sequences, their agonists, and

XX CC antagonists may have neuroprotective; cytostatic; cardioactive;

XX CC immunomodulatory; muscular active general; vulnerary; gastrointestinal

XX CC general; nephrotropic; antineoplastic; gynecological; or antibacterial

XX CC activity. The invention also includes antibodies specific for the

XX CC protein or polynucleotide sequences. The lung cancer associated

XX CC polynucleotide sequences may be used for detection of lung cancer,

XX CC chromosome identification, as chromosome markers, and for numerous other

XX CC diagnostic or research purposes. The proteins may be used to treat

XX CC disorders such as neural, immune, muscular, reproductive,

XX CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

XX CC disorders. The proteins may also be used in the treatment of wounds and

XX CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and

XX CC peptide AAF58549 are used in the course of the invention for the

XX CC identification and characterization of the polynucleotide and protein

XX CC sequences.

XX SQ Sequence 312 AA;

Query Match 50.4%; Score 645.5; DB 21; Length 312;

Best Local Similarity 58.0%; Pred. No. 1.5e-62;

Matches 123; Conservative 33; Mismatches 45; Indels 11; Gaps 3;

QY 42 FLEKFLVIGSAGTSGKSLHQLHOFIENKFKODSNHTIGVEGSRVNVGKTVKLIQIWDTA 101

DB 107 YIFKYIIIGDMGVGKSLHQLHOFIENKFKODSNHTIGVEGSRVNVGKTVKLIQIWDTA 166

QY 102 QERFRSVTRSYRGAAGALLVVDITSRTYNSLAWLTDARTLASPNIVVILCGNKKDL 161

DB 167 QERFRVTRSYRGAAGALLVVDITSRTYNSLAWLTDARTLASPNIVVILCGNKKDL 226

QY 162 PEREVTFLFASRFAQENELMFLTSALTGENVEAFLEAKKIQIYQIQSGDLNNAESG 221

DB 227 AQDYTYEAKQFAENGLLFLEASAKTGENVEDFALEAKKIQIYQIQSGDLNNAESG 286

QY 222 IQYGASLRQLRQPSRAQAV-AQP-----CGC 248

DB 287 VQKPSA-----POGGRUTSPQPOREGCGC 312.

## RESULT 14

ABB59884

ID ABB59884 standard; Protein; 215 AA.

XX ABB59884;

AC ABB59884;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 6444.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-658660/75.

DR N-PSDB; ABL03987.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.

PS Disclosure; SEQ ID NO 6444; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 215 AA;

Query Match 48.6%; Score 622.5; DB 22; Length 215;

Best Local Similarity 57.9%; Pred. No. 3.1e-60;

Matches 121; Conservative 29; Mismatches 56; Indels 3; Gaps 2;

QY 41 DELFKFLVIGSAGTSGKSLHQLHOFIENKFKODSNHTIGVEGSRVNVGKTVKLIQIWDTA 100

DB 9 NYIFKYIIIGDMGVGKSLHQLHOFIENKFKODSNHTIGVEGSRVNVGKTVKLIQIWDTA 68

QY 101 QERFRSVTRSYRGAAGALLVVDITSRTYNSLAWLTDARTLASPNIVVILCGNKKDL 160

DB 69 QERFRVTRSYRGAAGALLVVDITSRTYNSLAWLTDARTLASPNIVVILCGNKKDL 128

QY 161 PEREVTFLFASRFAQENELMFLTSALTGENVEAFLEAKKIQIYQIQSGDLNNAESG 220

DB 129 ESTREVTYEAKEAFENGLFLEASAKTGENVEDFALEAKKIQIYQIQSGDLNNAESG 188

QY 221 IQYGASLRQLRQPSRAQAV-AQP-----CGC 248

DB 189 VQKPSA-----POGGRUTSPQPOREGCGC 215

RESULT 15

ABG19365  
ID ABG19365 standard; Protein; 174 AA.

XX AC ABG19365;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #19356.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS83552.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX PS Claim 20; SEQ ID NO 49724; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG0010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence #174 AA;

Query Match 48.1%; Score 615.5; DB 22; Length 174;  
Best Local Similarity 79.7%; Pred. NO. 1.3e-59;  
Matches 122; Conservative 10; Mismatches 16; Indels 5; Gaps 1;

QY 96 IWDTAGQERFSVRSYRAGALLVYDITSREYNSLAAMLTARTLASPNIVILCG 155

DB 27 IWDTAGPDRFSRYSYHXRGLLVYDITAREYNTALTNLTDMRLASNVILCG 86

QY 156 NKKDLPDREVTFLASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDP 215

Db 87 NKKDLPDREVTFLASRFAQENELMFLTSALTGENVEEAFVQCARKILNKIES----- 141

QY 216 ERMGSGIQYGDASLRQLRQPRSAQAYAPQPCGC 248

Db 142 ERMGSGIQYGDAAALRLRSPRRRAQAPNAQECGC 174

Search completed: March 4, 2003, 14:58:50  
Job time : 23.3298 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: March 4, 2003, 14:53:52 ; Search time 11.1564 Seconds  
(without alignments)  
2137.006 Million cell updates/sec

Title: US-09-763-902B-8  
Perfect score: 1280  
Sequence: 1 MSVSLPLTVVRERDWIGIH.....LRQLRQPSAQAVAPQCGC 248

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

PIR\_73:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description         |
|------------|-------|-------------|--------|----------|---------------------|
| 1          | 1070  | 83.6        | 213    | 2 F36364 | GTP-binding protei  |
| 2          | 1065  | 83.2        | 213    | 2 S58279 | GTP-binding protei  |
| 3          | 938   | 73.3        | 213    | 2 E34323 | GTP-binding protei  |
| 4          | 937   | 73.2        | 213    | 2 S01766 | GTP-binding protei  |
| 5          | 645.5 | 50.4        | 215    | 2 T47160 | hypothetical protei |
| 6          | 640.5 | 50.0        | 215    | 2 E42148 | GTP-binding protei  |
| 7          | 606   | 47.3        | 210    | 2 T23530 | hypothetical protei |
| 8          | 556   | 43.4        | 213    | 2 S36367 | GTP-binding protei  |
| 9          | 556   | 43.4        | 213    | 2 J64106 | GTP-binding protei  |
| 10         | 550   | 43.0        | 211    | 2 S71585 | GTP-binding protei  |
| 11         | 544   | 42.5        | 210    | 2 T04362 | GTP-binding protei  |
| 12         | 543   | 42.4        | 212    | 2 B34323 | GTP-binding protei  |
| 13         | 540   | 42.2        | 209    | 2 T02242 | GTP-binding protei  |
| 14         | 539   | 42.1        | 212    | 2 A39648 | GTP-binding protei  |
| 15         | 539   | 42.1        | 212    | 2 S52325 | GTP-binding protei  |
| 16         | 539   | 42.1        | 212    | 2 S39979 | GTP-binding protei  |
| 17         | 539   | 42.1        | 214    | 2 T25796 | hypothetical protei |
| 18         | 538   | 42.0        | 212    | 2 S38341 | GTP-binding protei  |
| 19         | 537   | 42.0        | 223    | 2 T03767 | GTP-binding protei  |
| 20         | 536   | 41.9        | 202    | 2 B39663 | GTP-binding protei  |
| 21         | 535   | 41.8        | 210    | 2 T02248 | GTP-binding protei  |
| 22         | 534   | 41.7        | 208    | 2 E71440 | GTP-binding protei  |
| 23         | 534   | 41.7        | 211    | 2 H85191 | GTP-binding protei  |
| 24         | 533.5 | 41.7        | 234    | 2 T38030 | ras-related protei  |
| 25         | 527   | 41.2        | 211    | 2 S71559 | GTP-binding protei  |
| 26         | 503   | 39.3        | 205    | 2 D71440 | GTP-binding protei  |
| 27         | 461   | 36.0        | 223    | 2 S42679 | GTP-binding protei  |
| 28         | 455   | 35.5        | 211    | 2 T29035 | hypothetical protei |
| 29         | 453   | 35.4        | 222    | 2 T07059 | GTP-binding protei  |

30 449 35.1 224 2 T06105 GTP-binding protei  
31 448 35.0 225 2 T06445 GTP-binding protei  
32 447 34.9 218 2 C38625 GTP-binding protei  
33 447 34.9 218 2 A55005 GTP-binding protei  
34 445 34.8 218 2 J02487 GTP-binding protei  
35 443.5 34.6 214 2 T14566 GTP-binding protei  
36 443 34.6 216 2 JN0056 GTP-binding protei  
37 443 34.6 216 2 E36364 GTP-binding protei  
38 443 34.6 216 2 S47169 GTP-binding protei  
39 442.5 34.6 233 2 D96763 hypothetical protei  
40 442 34.5 222 2 T03622 GTP-binding protei  
41 441 34.5 203 2 T20750 hypothetical protei  
42 439 34.3 216 2 J04108 GTP-binding protei  
43 439 34.3 217 2 T12965 GTP-binding protei  
44 437.5 34.2 222 2 T06444 GTP-binding protei  
45 436 34.1 206 1 TVBYQ2 GTP-binding protei

## ALIGNMENTS

## RESULT 1

F36364  
GTP-binding protein rab4b - dog  
C:Species: Canis lupus familiaris (dog)  
C:Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 02-Feb-2001  
C:Accession: F36364; S15603  
R:Chavrier, P.; Vingron, M.; Sander, C.; Simons, K.; Zerial, M.  
Mol. Cell. Biol. 10, 6578-6585, 1990  
A:Title: Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell line.  
A:Reference number: A36364; MUID:91061765; PMID:2123294  
A:Accession: F36364  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-213 <CHA>  
A:Cross-references: GB:X56389; MID:g918; PIDN:CAA39800.1; PID:g919  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop  
F:9-124/Domain: translation elongation factor Tu homology <ETU>  
F:15-22/Region: nucleotide-binding motif A (P-loop)  
F:121-124/Region: GTP-binding NRXD motif  
F:151-153/Region: GTP-binding SAR/L motif  
F:211,213/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 83.6%; Score 1070; DB 2; Length 213;  
Best Local Similarity 100.0%; Pred. No. 1.4e-86;  
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 DFLFKFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGKTVKQLQIWDFA 100  
|||||

DB 6 DFLFKFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGKTVKQLQIWDFA 65  
|||||

QY 101 GQERFVSRTSYRGAAGALLVYDITSRETYNSLAALWDARTLASPNVILCGNKKDL 160  
|||||

DB 66 GQERFVSRTSYRGAAGALLVYDITSRETYNSLAALWDARTLASPNVILCGNKKDL 125  
|||||

QY 161 DPERVTFLEASRFAQENELMFLTSALTGENVEEFLKACARTILNKIDSGELDPERMGS 220  
|||||

DB 126 DPERVTFLEASRFAQENELMFLTSALTGENVEEFLKACARTILNKIDSGELDPERMGS 185  
|||||

QY 221 G1QYGDASLRQLRQPSAQAVAPQCGC 248  
|||||

DB 186 G1QYGDASLRQLRQPSAQAVAPQCGC 213  
|||||

## RESULT 2

S58279  
GTP-binding protein rab4b - rat  
N:Alternate names: ras-homolog GTPase  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 02-Feb-2001  
C:Accession: S58279  
R:Schuermann, A.; Muehl-Zuerbes, P.; Lie, C.; Joost, H.G.

DB 6 DFLPKFLVIGNACIGKSCLLHQFIEKFEKDDSNHTIGVEFGSKINVGKVKVQLINDTA 65

RESULT 5  
T47160  
hypothetical protein DKFzp762K0911.1 - human  
C:Species: Homo sapiens (man)

C; Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 02-Sep-2000  
C; Accession: T47160  
R; Blum, H.; Bauersachs, S.; Mewes, H.W.; Well, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000  
A:Reference number: Z24379

A:Accession: T47160

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-215 <AAA>

A:Cross-references: EMBL:AL162081

A:Experimental source: adult melanoma (Mewo cell line); clone DKF2p762K0911

C:Genetics:

A:Note: DKF2p762K0911.1

C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
F:12-127/Domain: translation elongation factor Tu homology <ETU>

Query Match 50.4%; Score 645.5; DB 2; Length 215;  
Best Local Similarity 58.0%; Pred. No. 2.8e-49;  
Matches 123; Conservative 33; Mismatches 45; Indels 11; Gaps 3;

QY 42 FLKFLVIGSAGTGKSCLLHOFIENKFKDSNHTTIGVEFGSRVYVNGGKTVKLQIWDTAG 101

Db 10 YIFKIIIGDMGVKSCLLHOFTEKKFMADCPHTIGVEFGTRIIEVSGKIKLQIWDTAG 69

QY 102 QERFSVTSYRGAAGALLYDITSRETYNSLAAWLTADARTLASPNIVILCGNKKDLD 161

Db 70 QERFVAVTSYRGAAGALMYDITRSTYNSHLSWLTADARTLASPNIVILCGNKKDLD 129

QY 162 PERVTFLASRFAOENELMFLTSALTGENVEAFKLCARTILNKIDSGELDPERMGG 221

Db 130 AQRDVTTEAKQFAEENGLLEASAKTGENVEDAFLEAAKIYQNIQDGLDNLNAESG 189

QY 222 IOYGDASLRQLRQPSQAQA-POP-----CGC 248

Db 190 VQHKPSA-----POGRLTSEPOQREGCGC 215

#### RESULT 6

E42148

GTP-binding protein rab14 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 02-Feb-2001

A:Accession: E42148

R:Elferink, L.A.; Anzai, K.; Scheller, R.H.

J. Biol. Chem. 267, 5768-5775, 1992

A:title: rab15, a novel low molecular weight GTP-binding protein specifically expressed

A:Reference number: E42148; MUID:92210533; PMID:1313420

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-215 <ELF>

A:Cross-references: GB:M83680

C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; lipoprotein; nucleotide binding; P-loop; prenylated cysteine

F:12-127/Domain: translation elongation factor Tu homology <ETU>

F:18-25/Region: nucleotide-binding motif A (P-loop)

F:124-127/Region: nucleotide-binding motif A (P-loop)

F:154-156/Region: GTP-binding NKXD motif

F:213,215/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 50.0%; Score 640.5; DB 2; Length 215;  
Best Local Similarity 58.0%; Pred. No. 7.7e-49;  
Matches 123; Conservative 32; Mismatches 46; Indels 11; Gaps 3;

QY 42 FLKFLVIGSAGTGKSCLLHOFIENKFKDSNHTTIGVEFGSRVYVNGGKTVKLQIWDTAG 101

Db 10 YIFKIIIGDMGVKSCLLHOFTEKKFMADCPHTIGVEFGTRIIEVSGKIKLQIWDTAG 69

QY 102 QERFSVTSYRGAAGALLYDITSRETYNSLAAWLTADARTLASPNIVILCGNKKDLD 161

Db 70 QERFVAVTSYRGAAGALMYDITRSTYNSHLSWLTADARTLASPNIVILCGNKKDLD 129

QY 162 PERVTFLASRFAOENELMFLTSALTGENVEAFKLCARTILNKIDSGELDPERMGG 221

Db 130 AQRDVTTEAKQFAEENGLLEASAKTGENVEDAFLEAAKIYQNIQDGLDNLNAESG 189

QY 222 IOYGDASLRQLRQPSQAQA-POP-----CGC 248

Db 190 VQHKPSA-----POGRLTSEPOQREGCGC 215

#### RESULT 7

T23530

hypothetical protein K09A9.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

A:Accession: T23530

R:Swinsburne, J.

submitted to the EMBL Data Library, August 1996

A:Reference number: 219753

A:Accession: T23530

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-210 <WIL>

A:Cross-references: EMBL:Z79601; PIDN:CAB01884.1; GSPDB:GNO0028; CESP:K09A9.2

A:Experimental source: clone K09A9

C:Genetics:

A:Gene: CESP:K09A9.2

A:Map position: X

A:Introns: 36/1; 117/3; 157/2

C:Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match 47.3%; Score 606; DB 2; Length 210;

Best Local Similarity 55.1%; Pred. No. 8.1e-46;

Matches 114; Conservative 32; Mismatches 55; Indels 6; Gaps 1;

QY 42 FLKFLVIGSAGTGKSCLLHOFIENKFKDSNHTTIGVEFGSRVYVNGGKTVKLQIWDTAG 101

Db 10 YIFKIIIGDMGVKSCLLHOFTEKKFMADCPHTIGVEFGTRIIEVSGKIKLQIWDTAG 69

QY 102 QERFSVTSYRGAAGALLYDITSRETYNSLAAWLTADARTLASPNIVILCGNKKDLD 161

Db 70 QERFVAVTSYRGAAGALMYDITRSTYNSHLSWLTADARTLASPNIVILCGNKKDLD 129

QY 162 PERVTFLASRFAOENELMFLTSALTGENVEAFKLCARTILNKIDSGELDPERMGG 221

Db 130 AQRDVTTEAKQFAEENGLLEASAKTGENVEDAFLEAAKIYQNIQDGLDNLNAESG 189

QY 222 IOYGDASLRQLRQPSQAQA-POP-----CGC 248

Db 190 VQ-----PKNLPRAAENNGKDCNC 210

#### RESULT 8

S36367

GTP-binding protein yptV4 - Volvox carteri

C:Species: Volvox carteri

C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 19-Jan-2001

A:Accession: S36367

R:Fabry, S.; Jacobsen, A.; Huber, H.; Palme, K.; Schmitt, R.

Curr. Genet. 24, 229-240, 1993

A:title: Structure, expression, and phylogenetic relationships of a family of ypt gen

A:Reference number: S36365; MUID:94037148; PMID:8221932

A:Accession: S36367

A:Molecule type: DNA

A:Residues: 1-213 <FAB>

A:Cross-references: GB:I08130; NID:g409165; PIDN:AAA34253.1; PID:g409166

C:Genetics:

A:Gene: yptV4

A:Introns: 4/2; 16/1; 62/3; 90/2

C:Superfamily: ras transforming protein; translation elongation factor Tu homology

C:Keywords: GTP binding; lipoprotein; nucleotide binding; P-loop

F:7-122/Domain: translation elongation factor Tu homology <ETU>

F:13-20/Region: nucleotide-binding motif A (P-loop)

F:59-66/Region: GTP binding #status predicted

F:116-123/Region: GTP binding #status predicted

F:145-151/Region: GTP binding #status predicted

Query Match 43.4%; Score 556; DB 2; Length 213;

C>Date: 06-Dec-1996 #sequence\_revision 27-Feb-1997 #text\_change 02-Feb-2001  
C:Accession: S71585; T04656; S71233  
R:Biermann, B.; Randall, S.K.; Crowell, D.N.  
Plant Mol. Biol. 31, 1021-1028, 1996  
A:Title: Identification and isoprenylation of plant GTP-binding proteins.  
A:Reference number: S71584; MUID:97000914; PMID:8843944  
A:Accession: S71585  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-211 <BEV>  
A:Cross-references: EMBL:U46925; NID:gl184982; PIDN:AA87883.1; PID:gl184983  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996  
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hohelsel, J.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, October 1998  
A:Reference number: Z15380  
A:Accession: T04696  
A:Molecule type: DNA  
A:Residues: 1-211 <BEV>  
A:Cross-references: EMBL:AL031986  
A:Experimental source: cultivar Columbia; BAC clone F4B14  
C:Genetics:  
A:Gene: ATGB2  
A:Map position: 4  
A:Introns: 16/1; 62/3; 90/2; 158/3; 181/3  
A:Note: F4B14.i30  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:7-122/Domain: translation elongation factor Tu homology <ETU>  
F:13-20/Region: nucleotide-binding motif A (P-loop)  
F:119-122/Region: GTP-binding NKXD motif

| Query Match           | 43.0%                                                           | Score 550;         | DB 2;     | Length 211; |
|-----------------------|-----------------------------------------------------------------|--------------------|-----------|-------------|
| Best Local Similarity | 58.4%                                                           | Pred. No. 6.7e-41; |           |             |
| Matches 108;          | Conservative 29;                                                | Mismatches 48;     | Indels 0; | Gaps 0;     |
| QY 41                 | DLFLFVLVGSAGTCKSCLLHOFIENKFKDQSNHTIGVEFGSRVNVVGGTKVLQIWDTA 100  |                    |           |             |
| Db 4                  | DYLFIYIIIGTGVGKSCLLLOFTDRFPQVHDLTIGVEFGARMTVDGRPIKLQIWDTA 63    |                    |           |             |
| QY 101                | GOERFSVTRSYRGAAGALLVYDITSRTYNSLAAWLTADARTLASPNVIVILCGNKKDL 160  |                    |           |             |
| Db 64                 | GOESFSITRSYRGAAGALLVYDITRFTFNHLASWLEDARQHANPNMSIMLIGNKKDL 123   |                    |           |             |
| QY 161                | DPREYFTFLASFAQENELMFLTETSALTGENVEEAFILKARTILNKIDSGELDPERMGS 220 |                    |           |             |
| Db 124                | AHKRAYSKKEGQCFAXEHGLLFLASARTAQNVEEAFIETAAKILQIQGVDFVDSNESS 183  |                    |           |             |
| QY 221                | GIOYG 225                                                       |                    |           |             |
| Db 184                | GIKIG 188                                                       |                    |           |             |

RESULT 11  
T04362  
GTP-binding protein yptm3 - maize  
N:Alternate names: ras-like protein yptm3  
C:Species: Zea mays (maize)  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 02-Feb-2001  
R:Diefenthal, T.  
submitted to the EMBL Data Library, February, 1994  
A:Reference number: Z15312  
A:Accession: T04362  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-210 <DIE>  
A:Cross-references: EMBL:X77795; PIDN:CAA54822.1  
C:Genetics:  
A:Gene: yptm3  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; membrane trafficking; nucleotide binding; P-loop  
F:7-122/Domain: translation elongation factor Tu homology <ETU>  
F:13-20/Region: nucleotide-binding motif A (P-loop)

A;Status: preliminary



A:Molecule type: mRNA  
A:Residues: 1-38,'M',40-212 <CHW>  
C:Genetics:

A:Gene: rab2  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;  
F:7-122/Domain: translation elongation factor Tu homology <ETU>  
F:13-20/Region: nucleotide-binding motif A (P-loop)  
F:119-122/Region: GTP-binding NAKD motif  
F:149-151/Region: GTP-binding SAK/L motif  
F:211,212/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 42.1%; Score 539; DB 2; Length 212;  
Best Local Similarity 56.0%; Pred. No. 6.3e-40;  
Matches 103; Conservative 30; Mismatches 51; Indels 0; Gaps 0;

QY 42 FLPLFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGGKTVKQIWDITAG 101  
DB 5 YLFKXIIIGDTGVGKSCLLLOFTDKRPQVHDLTIGVEFGARMITIDGKQIKQIWDITAG 64  
QY 102 QERFSVTSRYRGAAGALLVYDITRETYNSLAAMLTDARTLASPNVIVILCGNKKDLD 161  
DB 65 QESFRSITRSYRGAAGALLVYDITRDTFNHLLTWLEDAHQHSNNVMVIMLIGNKSDLE 124  
QY 162 PERVTFLASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
DB 125 SRREVKEEGEAFAREHGLIFMETSAKTSNVVEAFINTAKEIYEKIQEGVFDINNEANG 184  
QY 222 IQYG 225  
DB 185 IKIG 188

RESULT 15  
S52325  
GTP-binding protein RAB2 - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 08-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 02-Feb-2001  
C:Accession: S52325  
R:Montpellier, C.; Kherrouche, Z.; Beque, A.; Stehelin, D.; Coll, J.  
submitted to the EMBL Data Library, January 1995  
A:Description: RAB2 nucleotide sequence in Gallus gallus and its phylogenetic position.  
A:Reference number: S52325  
A:Accession: S52325  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-212 <MON>  
A:Cross-references: EMBL:X84220; NID:G664924; PIDN:CAA59004.1; PID:G664925  
C:Genetics:

A:Gene: RAB2  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;  
F:7-122/Domain: translation elongation factor Tu homology <ETU>  
F:13-20/Region: nucleotide-binding motif A (P-loop)  
F:119-122/Region: GTP-binding NAKD motif  
F:149-151/Region: GTP-binding SAK/L motif  
F:211,212/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 42.1%; Score 539; DB 2; Length 212;  
Best Local Similarity 56.0%; Pred. No. 6.3e-40;  
Matches 103; Conservative 30; Mismatches 51; Indels 0; Gaps 0;

QY 42 FLPLFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGGKTVKQIWDITAG 101  
DB 5 YLFKXIIIGDTGVGKSCLLLOFTDKRPQVHDLTIGVEFGARMITIDGKQIKQIWDITAG 64  
QY 102 QERFSVTSRYRGAAGALLVYDITRETYNSLAAMLTDARTLASPNVIVILCGNKKDLD 161  
DB 65 QESFRSITRSYRGAAGALLVYDITRDTFNHLLTWLEDAHQHSNNVMVIMLIGNKSDLE 124  
QY 162 PERVTFLASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
DB 125 SRREVKEEGEAFAREHGLIFMETSAKTSNVVEAFINTAKEIYEKIQEGVFDINNEANG 184

QY 222 IQYG 225  
DB 185 IKIG 188

Search completed: March 4, 2003, 15:04:52  
Job time: 11.1564 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw\_model

Run on: March 4, 2003, 14:51:02 ; Search time 5.93644 Seconds  
(without alignments)  
1732.709 Million cell updates/sec

Title: US-09-763-902B-8

Perfect score: 1280

Sequence: 1 MSVSLPLTVMYRERDWIGIH.....LRLQRPRAQAVAPQPCGC 248

Scoring table: BLOSUM62

Gapop: 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match % | Length | ID           | Description        |
|------------|-------|---------------|--------|--------------|--------------------|
| 1          | 1070  | 83.6          | 213    | 1 RB4B_HUMAN | P22750 homo sapien |
| 2          | 1065  | 83.2          | 213    | 1 RB4B_RAT   | P51146 rattus norv |
| 3          | 942   | 73.6          | 213    | 1 RB4A_MOUSE | P56371 mus musculu |
| 4          | 938   | 73.3          | 213    | 1 RB4A_HUMAN | P20338 homo sapien |
| 5          | 937   | 73.2          | 213    | 1 RB4A_RAT   | P05714 rattus norv |
| 6          | 645.5 | 50.4          | 215    | 1 RB14_HUMAN | P35287 homo sapien |
| 7          | 623   | 48.7          | 205    | 1 RB4A_DICDI | P36410 dictyosteli |
| 8          | 556   | 43.4          | 213    | 1 YPT4_CHLRE | Q39570 chlamydomon |
| 9          | 556   | 43.4          | 213    | 1 YPT4_VOLCA | P36863 volvox cart |
| 10         | 540   | 42.2          | 209    | 1 RB2A_WAIZE | P49103 zea mays (m |
| 11         | 539   | 42.1          | 212    | 1 RB2B_HUMAN | P08886 homo sapien |
| 12         | 539   | 42.1          | 212    | 1 RB2B_MOUSE | P53994 mus musculu |
| 13         | 539   | 42.1          | 212    | 1 RB2B_RABIT | Q01971 oryctolagus |
| 14         | 538   | 42.0          | 212    | 1 RB2B_LYMS  | Q05975 lymnaea sta |
| 15         | 536   | 41.9          | 212    | 1 RB2B_RAT   | P05712 rattus norv |
| 16         | 535   | 41.8          | 210    | 1 RB2B_WAIZE | P49104 zea mays (m |
| 17         | 533.5 | 41.7          | 234    | 1 YE1B_SCHPO | O13876 schizosacch |
| 18         | 524   | 40.9          | 199    | 1 RB2B_DICDI | P36409 dictyosteli |
| 19         | 461   | 36.0          | 223    | 1 YP13_YEAST | P38555 saccharomyc |
| 20         | 447   | 34.9          | 218    | 1 R11B_DISOM | P22129 discopyge o |
| 21         | 447   | 34.9          | 218    | 1 R11B_MOUSE | P46638 mus musculu |
| 22         | 445   | 34.8          | 218    | 1 R11B_HUMAN | Q15907 homo sapien |
| 23         | 443.5 | 34.6          | 214    | 1 RB2B_BETU  | Q39434 beta vulgar |
| 24         | 443   | 34.6          | 216    | 1 R11A_HUMAN | P24410 homo sapien |
| 25         | 443   | 34.6          | 216    | 1 R11A_MOUSE | Q911x1 mus musculu |
| 26         | 443   | 34.6          | 226    | 1 R11A_LOTJA | Q40191 lotus japon |
| 27         | 442   | 34.5          | 222    | 1 R11D_TOBAC | Q40522 nicotiana t |
| 28         | 439   | 34.3          | 216    | 1 YPT6_CHLRE | Q39572 chlamydomon |
| 29         | 439   | 34.3          | 217    | 1 RB1A_ARATH | Q96283 arabidopsis |
| 30         | 436   | 34.1          | 206    | 1 YPT1_YEAST | P01123 saccharomyc |
| 31         | 436   | 34.1          | 216    | 1 R11C_LOTJA | Q40193 lotus japon |
| 32         | 435   | 34.0          | 222    | 1 R11C_TOBAC | Q40520 nicotiana t |
| 33         | 433   | 33.8          | 203    | 1 YPT1_VOLCA | P31584 volvox cart |

34 432 33.8 218 1 R11B\_RAT O35509 rattus norv  
35 431.5 33.7 205 1 RB1B\_LYMS O05974 lymnaea sta  
36 431 33.7 217 1 RB1C\_ARATH O04486 arabidopsis  
37 431 33.7 218 1 R11D\_LOTJA Q40194 lotus japon  
38 430.5 33.6 216 1 RB1B\_ARATH Q39222 arabidopsis  
39 430 33.6 214 1 RB11\_DICDI P36412 dictyosteli  
40 430 33.6 214 1 YPT3\_SCHPO P17610 schizosacch  
41 429.5 33.6 203 1 YPT1\_CHLRE Q39571 chlamydomon  
42 428.5 33.5 213 1 RB25\_MOUSE Q9wtl12 mus musculu  
43 428.5 33.5 216 1 ARA2\_ARATH P28185 arabidopsis  
44 428.5 33.5 222 1 YP32\_YEAST P51996 saccharomyc  
45 426.5 33.3 217 1 RGP2\_ORYSA Q40723 oryza sativ

## ALIGNMENTS

RESULT 1  
RB4B\_HUMAN  
ID RB4B\_HUMAN STANDARD; PRT; 213 AA.  
AC P22750;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ras-related protein Rab-4B.  
GN RAB4B.  
OS Homo sapiens (Human), and  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606, 9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human.  
RA Huang C., Wu T., Xu S., Gu W., Wang Y., Han Z., Chen Z.;  
RT "Novel genes expressed in hematopoietic stem/progenitor cells from  
RT myelodysplastic syndromes patient."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=C.familiaris; STRAIN=Cocker spaniel;  
RX MEDLINE=91061765; PubMed=2123294;  
RA Chavrier P., Vingron M., Sander C., Simons K., Zerial M.;  
RT "Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell  
RT line."  
RL Mol. Cell. Biol. 10:6578-6585(1990).  
CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
CC TRAFFIC (BY SIMILARITY).  
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; AF165522; AAD45923.1;  
CC EMBL; X56389; CAA39800.1;  
CC PIR; F36364; F36364.  
CC HSSP; P36017; LEKO.  
CC Genew; HGNC:9782; RAB4B.  
CC InterPro; IPR003579; GTPase\_Rab.  
CC InterPro; IPR001806; Ras\_trnsmfmg.  
CC InterPro; IPR005225; Small\_GTP.  
CC Pfam; PF00071; ras.1.  
CC PRINTS; PR00449; RASTENSRFMNG.  
CC SMART; SM00175; RAB; 1.  
CC TIGRFAMs; TIGR00231; small\_GTP; 1.  
KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
FT

FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
 FT LIPID 211 211 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 213 213 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 213 AA; 23587 MW; 0C3D76DC3285DB98 CRC64;

Query Match 83.6%; Score 1070; DB 1; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-83;  
 Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 DFLFKFLVIGSAGTGKSCLLHOFIENFKQDSNHTIGVEFGSRVNVVGGKTKVLIQIWDTA 100  
 DB 6 DFLFKFLVIGSAGTGKSCLLHOFIENFKQDSNHTIGVEFGSRVNVVGGKTKVLIQIWDTA 65  
 QY 101 GOEFRSRTSYRYGAAGALLVYDITSRETYNSLAAMLTDTARTLASPNIVVILCGNKKDL 160  
 DB 66 GOEFRSRTSYRYGAAGALLVYDITSRETYNSLAAMLTDTARTLASPNIVVILCGNKKDL 125  
 QY 161 DPEREVTFLEASRAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMG 220  
 DB 126 DPEREVTFLEASRAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMG 185  
 QY 221 GIQYGDASLRQLRPSRAQAVAPQCGC 248  
 DB 186 GIQYGDASLRQLRPSRAQAVAPQCGC 213

RESULT 2  
 RB4B\_RAT STANDARD; PRT; 213 AA.  
 ID RB4B\_RAT  
 AC P51146;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE Ras-related protein Rab-4B.  
 GN RAB4B.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-Sprague-Dawley; TISSUE-Heart muscle;  
 RA Schuermann A., Muehl-Zuerbes P., Lie C., Joost H.G.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC -1- TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X78605; CAA5339.1;  
 DR HSSP; P36017; 1EKO.  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsfrmg.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTRNSFRMG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
 FT LIPID 211 211 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 213 213 GERANYL-GERANYL (BY SIMILARITY).

SQ SEQUENCE 213 AA; 23629 MW; 0C3D76DC328B0018 CRC64;  
 Query Match 83.2%; Score 1065; DB 1; Length 213;  
 Best Local Similarity 99.5%; Pred. No. 6.1e-83;  
 Matches 207; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 41 DFLFKFLVIGSAGTGKSCLLHOFIENFKQDSNHTIGVEFGSRVNVVGGKTKVLIQIWDTA 100  
 DB 6 DFLFKFLVIGSAGTGKSCLLHOFIENFKQDSNHTIGVEFGSRVNVVGGKTKVLIQIWDTA 65  
 QY 101 GOEFRSRTSYRYGAAGALLVYDITSRETYNSLAAMLTDTARTLASPNIVVILCGNKKDL 160  
 DB 66 GOEFRSRTSYRYGAAGALLVYDITSRETYNSLAAMLTDTARTLASPNIVVILCGNKKDL 125  
 QY 161 DPEREVTFLEASRAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMG 220  
 DB 126 DPEREVTFLEASRAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMG 185  
 QY 221 GIQYGDASLRQLRPSRAQAVAPQCGC 248  
 DB 186 GIQYGDASLRQLRPSRAQAVAPQCGC 213

RESULT 3  
 RB4A\_MOUSE STANDARD; PRT; 213 AA.  
 ID RB4A\_MOUSE  
 AC P56371;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DE Ras-related protein Rab-4A.  
 GN RAB4A OR RAB4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE-Liver;  
 RX MEDLINE=97107274; PubMed=8950022;  
 RA Ikeda H., Ikegami T., Mitsui T., Senda D., Hayasaka K.;  
 RT "Isolation and sequence determination of cDNA encoding mouse rab 4  
 RT and candidate approach for the beige mutation in mice."  
 RL Biochem. Mol. Biol. Int. 40:647-651(1996).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC -1- TRAFFIC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: GENERALLY ASSOCIATED WITH MEMBRANES;  
 CC CYTOPLASMIC WHEN PHOSPHORYLATED BY CDC2 (BY SIMILARITY).  
 CC -1- PTM: PHOSPHORYLATED BY CDC2-KINASE DURING MITOSIS (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; D86563; BAA24034.1;  
 DR HSSP; P05713; 3RAB.  
 DR MGD; MGI:105069; Rab4a.  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsfrmg.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTRNSFRMG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport;  
 KW Phosphorylation.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).

FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
 FT MOD\_RES 199 199 PHOSPHORYLATION (BY CDC2)  
 (BY SIMILARITY).  
 FT LIPID 211 211 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 213 213 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 213 AA; 23938 MW; 9FEF188966922AD6 CRC64;

Query Match 73.6%; Score 942; DB 1; Length 213;  
 Best Local Similarity 86.5%; Pred. No. 1.5e-72;  
 Matches 180; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

OY 41 DFLFKFLVIGSAGTCKSLHOFIENKFKQDSNHHITGVEFGSRVNVGGKTVKQLQIWDTA 100  
 DB 6 DFLFKFLVIGNAGTCKSLHOFIENKFKQDSNHHITGVEFGSKIIINVGGKYVKQLQIWDTA 65  
 OY 101 GOERFSRVSRYRGAAGALLVDITSRTYNSLAALWLTDAFTLASPNVIVILCGNKKDL 160  
 DB 66 GOERFSRVSRYRGAAGALLVDITSRTYNSLAALWLTDAFTLASPNVIVILCGNKKDL 125  
 OY 161 DPERVTFLEASRFAQENELMFLTSALTGENVEEAFKCAITLKNKIDSGELDPERMGS 220  
 DB 126 DADREVTFLASRFAQENELMFLTSALTGENVEEAFWQCAKILNKIESGELDPERMGS 185  
 OY 221 GYOYGDASLRQLRSPRSQAQVAPQCGC 248  
 DB 186 GYOYGDALRLRSPRTPQAQEGCG 213

RESULT 4  
 ID RB4A\_HUMAN STANDARD; PRT; 213 AA.  
 AC P20338;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Ras-related protein Rab-4A.  
 GN RAB4A OR RAB4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89308668; PubMed=2501306;  
 RA Zahraoui A., Touchot N., Chardin P., Tavittian A.;  
 RT "The human Rab genes encode a family of GTP-binding proteins related  
 to yeast YPT1 and SEC4 products involved in secretion.";  
 RL J. Biol. Chem. 264:12394-12401(1989).  
 RN [2]  
 RP PHOSPHORYLATION BY CDC2.  
 RX MEDLINE=91218852; PubMed=1902553;  
 RA Bailly E., McCaffrey M., Touchot N., Zahraoui A., Goud B., Bornens M.;  
 RT "Phosphorylation of two small GTP-binding proteins of the Rab family  
 by p34cdc2".  
 RL Nature 350:715-718(1991).  
 RN [3]  
 RP PHOSPHORYLATION BY CDC2.  
 RX MEDLINE=93049188; PubMed=1425574;  
 RA van der Sluis P., Hull M., Huber L.A., Male P., Goud B., Mellman I.;  
 RT "Reversible phosphorylation-dephosphorylation determines the  
 localization of rab4 during the cell cycle.";  
 RL EMBO J. 11:4379-4382(1992).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 TRAFFIC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: GENERALLY ASSOCIATED WITH MEMBRANES;  
 CYTOPLASMIC WHEN PHOSPHORYLATED BY CDC2.  
 CC -1- PTM: PHOSPHORYLATED BY CDC2-KINASE DURING MITOSIS.  
 CC -1- SIMILARITY: TO RAB PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; M28211; AAA60244.1;  
 DR PIR; E34323; E34323.  
 DR HSSP; P36017; 1EK0.  
 DR Genew; HGNC:9781; RAB4A.  
 DR MIM; 179511;  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsfrmg.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras. 1.  
 DR PRINTS; PR00449; RASTRNSFRMNG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRFAMS; TIGR00231; small\_gtp. 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport;  
 KW Phosphorylation.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
 FT MOD\_RES 199 199 PHOSPHORYLATION (BY CDC2).  
 FT LIPID 211 211 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 213 213 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 213 AA; 23902 MW; C4F20489D148FAE1 CRC64;

Query Match 73.3%; Score 938; DB 1; Length 213;  
 Best Local Similarity 86.5%; Pred. No. 3.2e-72;  
 Matches 180; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

OY 41 DFLFKFLVIGSAGTCKSLHOFIENKFKQDSNHHITGVEFGSRVNVGGKTVKQLQIWDTA 100  
 DB 6 DFLFKFLVIGNAGTCKSLHOFIENKFKQDSNHHITGVEFGSKIIINVGGKYVKQLQIWDTA 65  
 OY 101 GOERFSRVSRYRGAAGALLVDITSRTYNSLAALWLTDAFTLASPNVIVILCGNKKDL 160  
 DB 66 GOERFSRVSRYRGAAGALLVDITSRTYNSLAALWLTDAFTLASPNVIVILCGNKKDL 125  
 OY 161 DPERVTFLEASRFAQENELMFLTSALTGENVEEAFKCAITLKNKIDSGELDPERMGS 220  
 DB 126 DADREVTFLASRFAQENELMFLTSALTGENVEEAFVQCAKILNKIESGELDPERMGS 185  
 OY 221 GYOYGDASLRQLRSPRSQAQVAPQCGC 248  
 DB 186 GYOYGDALRLRSPRTPQAQEGCG 213

RESULT 5  
 ID RB4A\_RAT STANDARD; PRT; 213 AA.  
 AC P05714;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Ras-related protein Rab-4A.  
 GN RAB4A OR RAB4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88068563; PubMed=3317403;  
 RA Touchot N., Chardin P., Tavittian A.;  
 RT "Four additional members of the ras gene superfamily isolated by an  
 oligonucleotide strategy: molecular cloning of YPT-related cDNAs from  
 a rat brain library.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8210-8214(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;

RX MEDLINE-88143997; PubMed-3344209;  
 RA Zahroui A., Touchot N., Chardin P., Tavittian A.;  
 RT "Complete coding sequences of the ras related rab 3 and 4 cDNAs.";   
 RL Nucleic Acids Res. 16:1204-1204(1988).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: GENERALLY ASSOCIATED WITH MEMBRANES;  
 CC CYTOPLASMIC WHEN PHOSPHORYLATED BY CDC2 (BY SIMILARITY).  
 CC -1- PTM: PHOSPHORYLATED BY CDC2-KINASE DURING MITOSIS (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X06890; CAA30006.1;  
 DR PIR; S01766; S01766.  
 DR PIR; D39963; D39963.  
 DR HSSP; P36017; LEKO.  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsmg.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASRNSFRMG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport;  
 KW Phosphorylation.  
 FT NP\_BIND 35 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
 FT MOD\_RES 199 199 PHOSPHORYLATION (BY CDC2)  
 FT (BY SIMILARITY).  
 FT LIPID 211 211 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 213 213 GERANYL-GERANYL (BY SIMILARITY).  
 FT CONFLICT 152 152 A -> P (IN REF. 2).  
 SQ SEQUENCE 213 AA; 23892 MW; 34735DFCFE6470CF CRC64;  
  
 Query Match 73.2%; Score 937; DB 1; Length 213;  
 Best Local Similarity 86.1%; Pred. No. 3.9e-72;  
 Matches 179; Conservative 12; Mismatches 17; Indels 0; Gaps 0;  
  
 QY 41 DFLFKFLVIGSAGTGSCLLHQFIENKFKDSDNHTIGVEFGSRVNVGKTVKLQIWDTA 100  
 DB 6 DFLFKFLVIGSAGTGSCLLHQFIENKFKDSDNHTIGVEFGKLIINVGKTVKLQIWDTA 65  
  
 QY 101 GQERFRSVTSYRGAAGALLVYDITSRETYNSLAALWLTARTLASPNVIVILCGNKKDL 160  
 DB 66 GQERFRSVTSYRGAAGALLVYDITSRETYNLTNLTDLARMLASQNIIVILCGNKKDL 125  
  
 QY 161 DPREVTFLEASFAQENELMFLTSALTGCVNEEAFKCARILNKIDSGELDPERMG 220  
 DB 126 DAREVTFLEASFAQENELMFLTSALTGCVNEEAFKCARILNKIDSGELDPERMG 185  
  
 QY 221 GIOYGDASLRQLRPSRAQAVAPQCC 248  
 DB 186 GIOYGDASLRQLRPSRAQAVAPQCC 213  
  
 RESULT 6  
 RB14\_HUMAN STANDARD; PRT; 215 AA.  
 ID AC P35287; Q9U11; Q969L0;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 15-FEB-1994 (Rel. 28, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ras-related protein Rab-14.  
 GN RAB14.

OS Homo sapiens (Human), and  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606, 10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human; TISSUE=Liver;  
 RA Proikas-Cezanne T., Jenkins J.R.;  
 RT "Human Rab14 cloning and intracellular localization to the  
 RL biosynthetic/secretory trafficking pathway.";   
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human;  
 RA Ren Y.;  
 RT "Cloning and characterization of human small GTPase Rab14.";   
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human; TISSUE=Adrenal gland;  
 RX MEDLINE-20402571; PubMed-10931946;  
 RA Hu R.-M., Han Z.-G., Song H.-D., Peng Y.-D., Huang Q.-H., Ren S.-X.,  
 RA Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H.,  
 RA Gu B.-W., Dai M., Mao Y.-F., Gao G.-F., Rong R., Ye M., Zhou J.,  
 RA Xu S.-H., Gu J., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,  
 RA Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.;  
 RT "Gene expression profiling in the human hypothalamus-pituitary-adrenal  
 RL axis and full-length cDNA cloning.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human; TISSUE=Melanoma;  
 RA Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human; TISSUE=Placenta;  
 RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;  
 RT "NEDO human cDNA sequencing project.";   
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human;  
 RA Laird G.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human; TISSUE=Placenta;  
 RA Strausberg R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Rat; STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RX MEDLINE-92210533; PubMed-1313420;  
 RA Elferink L.A., Anzai K., Scheller R.H.;  
 RT "Rab15, a novel low molecular weight GTP-binding protein specifically  
 RL expressed in rat brain.";   
 RL J. Biol. Chem. 267:5768-5775(1992).  
 CC -1- FUNCTION: MAY BE INVOLVED IN VESICULAR TRAFFICKING AND  
 CC NEUROTRANSMITTER RELEASE.  
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN WHOLE BRAIN, SPINAL  
 CC CORD, HEART, KIDNEY AND LUNG.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

| RESULT 7 | RAB4_DICD1                                                          | STANDARD; | PRT; | 205 AA. |
|----------|---------------------------------------------------------------------|-----------|------|---------|
| ID       | RAB4_DICD1                                                          |           |      |         |
| AC       | P36410;                                                             |           |      |         |
| DT       | 01-JUN-1994 (Rel. 29, Created)                                      |           |      |         |
| DT       | 01-JUN-1994 (Rel. 29, Last sequence update)                         |           |      |         |
| DT       | 01-NOV-1997 (Rel. 35, Last annotation update)                       |           |      |         |
| DE       | Ras-related protein Rab4.                                           |           |      |         |
| GN       | RABD OR RAB4.                                                       |           |      |         |
| OS       | Dictyostelium discoideum (Slime mold).                              |           |      |         |
| OC       | Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.                |           |      |         |
| OX       | NCBI_TaxID=44689;                                                   |           |      |         |
| RN       | [1]                                                                 |           |      |         |
| RP       | SEQUENCE FROM N. A.                                                 |           |      |         |
| RC       | STRAIN=AX3;                                                         |           |      |         |
| RX       | MEDLINE=95181582; PubMed=7876348;                                   |           |      |         |
| RA       | Bush J.M. IV, Nolte K., Rodriguez-Paris J., Kaufmann N.,            |           |      |         |
| RA       | O'Halloran T., Ruscetti T., Ternesvari L., Steck T., Cardelli J.A.; |           |      |         |

```
CC CC TRAFFIC (BY SIMILARITY).
CC CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U13167; AAA82726.1;
CC CC HSPF; P05713; 3RAB.
CC CC InterPro; IPR003579; GTPase_Rab.
CC CC InterPro; IPR001806; Ras_trnsmg.
CC CC Pfam; PF00071; ras; 1.
CC CC PRINTS; PR00449; RASTRNSFRMG.
CC CC SMART; SM00175; RAB; 1.
CC CC TIGRfams; TIGR00231; small_GTP; 1.
CC CC KW GTP-binding; Lipoprotein; Prenylation; Protein transport;
CC CC Multigene family.
CC CC FT NP_BIND 13 20 GTP (BY SIMILARITY).
CC CC FT NP_BIND 61 65 GTP (BY SIMILARITY).
CC CC FT NP_BIND 119 122 GTP (BY SIMILARITY).
CC CC FT DOMAIN 35 43 EFFECTOR REGION (PROBABLE).
CC CC FT LIPID 212 212 GERANYL-GERANYL (BY SIMILARITY).
CC CC FT LIPID 213 213 GERANYL-GERANYL (BY SIMILARITY).
CC CC FT LIPID 213 213 GERANYL-GERANYL (BY SIMILARITY).
CC CC SQ SEQUENCE 213 AA; 23598 MW; CB9B3AAAE4E8BA76 CRC64;

Query Match 43.4%; Score 556; DB 1; Length 213;
Best Local Similarity 58.7%; Pred. No. 5.6e-40;
Matches 108; Conservative 27; Mismatches 49; Indels 0; Gaps 0;

QY 42 FLFKFLVLSAGTGKSCLLHQFTENKFKQDSNHTIGVEGSRVNVGGTKVQLQIWDTAG 101
Db 5 YLFKYIIIGDTGVGKSCLLQLQFTDKRFQPHDLTIGVEGARMINDGKQIKLIWDTAG 64

QY 102 QERFRSVTSYRGAAGALLVYDITRETYNSLAAMLTARTLASPNVIVILGCKNKDLD 161
Db 65 QESFRSITRYSYRGAAGALLVYDITRETFNHLASLEDAHQHANPNMTIMLGKCDLT 124

QY 162 PERVTFLASRAEQNELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221
Db 125 HRAVTEGEQFAKEGHLIFLETSARTAHNVEEAFINTAKEIYKKIQDGVFVDSNESY 184

QY 222 IQYG 225
Db 185 IKVG 188

RESULT 9
ID YPT4_VOLCA STANDARD; PRT; 213 AA.
AC P36863;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE GTP-binding protein ypt4 (RAB2 homolog).
GN YPTV4.
OS Volvox carteri.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OX NCBI_TaxID=3067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=f. Nagariensis / HK10;
RX MEDLINE=34037148; PubMed=8221932;
RA Fabry S., Jacobsen A., Huber H., Palme K., Schmitt R.;
RT "Structure, expression, and phylogenetic relationships of a family of
RT ypt genes encoding small G-proteins in the green alga Volvox
RT carteri.";
RL Curr. Genet. 24:229-240(1993).
```

```
CC CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC CC TRAFFIC (BY SIMILARITY).
CC CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; L08130; AAA34253.1;
CC CC PIR; S36367; S36367.
CC CC HSPF; P05713; 3RAB.
CC CC InterPro; IPR003579; GTPase_Rab.
CC CC InterPro; IPR001806; Ras_trnsmg.
CC CC Pfam; PF00071; ras; 1.
CC CC PRINTS; PR00449; RASTRNSFRMG.
CC CC SMART; SM00175; RAB; 1.
CC CC TIGRfams; TIGR00231; small_GTP; 1.
CC CC KW GTP-binding; Lipoprotein; Prenylation; Protein transport;
CC CC Multigene family.
CC CC FT NP_BIND 13 20 GTP (BY SIMILARITY).
CC CC FT NP_BIND 61 65 GTP (BY SIMILARITY).
CC CC FT NP_BIND 119 122 GTP (BY SIMILARITY).
CC CC FT DOMAIN 35 43 EFFECTOR REGION (PROBABLE).
CC CC FT LIPID 212 212 GERANYL-GERANYL (BY SIMILARITY).
CC CC FT LIPID 213 213 GERANYL-GERANYL (BY SIMILARITY).
CC CC FT LIPID 213 213 GERANYL-GERANYL (BY SIMILARITY).
CC CC SQ SEQUENCE 213 AA; 23697 MW; FFE43AAAE4E8B5E6 CRC64;

Query Match 43.4%; Score 556; DB 1; Length 213;
Best Local Similarity 58.7%; Pred. No. 5.6e-40;
Matches 108; Conservative 27; Mismatches 49; Indels 0; Gaps 0;

QY 42 FLFKFLVLSAGTGKSCLLHQFTENKFKQDSNHTIGVEGSRVNVGGTKVQLQIWDTAG 101
Db 5 YLFKYIIIGDTGVGKSCLLQLQFTDKRFQPHDLTIGVEGARMINDGKQIKLIWDTAG 64

QY 102 QERFRSVTSYRGAAGALLVYDITRETYNSLAAMLTARTLASPNVIVILGCKNKDLD 161
Db 65 QESFRSITRYSYRGAAGALLVYDITRETFNHLASLEDAHQHANPNMTIMLGKCDLT 124

QY 162 PERVTFLASRAEQNELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221
Db 125 HRAVTEGEQFAKEGHLIFLETSARTAHNVEEAFINTAKEIYKKIQDGVFVDSNESY 184

QY 222 IQYG 225
Db 185 IKVG 188

RESULT 10
ID RB2A_MAIZE STANDARD; PRT; 209 AA.
AC P49103;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Ras-related protein Rab-2-A.
GN RB2A.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Laughner B.J., Ferl R.J., Almira E.C.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC TRAFFIC (BY SIMILARITY).
```





```

FT LIPID 212 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 212 AA; 23547 MW; 58231E3F9EF007A5 CRC64;

Query Match 42.1%; Score 539; DB 1; Length 212;
Best Local Similarity 56.0%; Pred: No. 1.5e-38;
Matches 103; Conservative 30; Mismatches 51; Indels 0; Gaps 0;

QY 42 FLKFLVIGSAGTCKSCLLHQFIENKFQODSNHPTGVFEFSNVNNGVKTVKLQIWDTAG 101
 :|:|:|:|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5 YLFKYIIIGTCVGKSCLLQLQFTDKRPVPHDLTIGVEFGARMITDGKIKLOIWDTAG 64

QY 102 QERFSRVSRYRGAAGALLYDYDTSRETYNSLAAWLTDAATLASPNVTWICGNKKOLD 161
 || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db .65 QESERSIPRSYRGAAGALLYDYDIRRTDFNHLTWLEDAQRQHSSNNVMILGNKSdle 124.

QY 162 PEREVTFLEASRFQENELMFLETSAITGENVEEAFPLKCAFTILNKIDSGELDPRMGSG 221
 || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 125 SRREVVKEGEFAAREHGLIFMETSAKTASNVEEAFINTAKEIYEKIQEGVPDIINNEANG 184

QY 222 IOYG 225
 | | | | |
Db 185 IKIG 188

```

RESULT 13  
RAB2\_RAB1T  
ID RAB2\_RAB1T STANDARD; PRT; 212 AA.  
AC Q01971;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE Ras-related protein Rab-2.  
GN RAB2.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OC NCBI\_TaxID=9986;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=93359934; PubMed=1323270;  
RA TANG L.H., Stoch S.A., Modlin I.M., Goldenring J.R.;  
RX Identification of rab2 as a tubulovesicle-membrane-associated  
RT protein in rabbit gastric parietal cells.;  
RT Biochem. J. 285:715-719(1992).  
RL  
CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
CC -1- TRAFFIC (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH A STRUCTURE HAVING THE  
CC CHARACTERISTICS OF AN IMMEDIATE COMPARTMENT BETWEEN THE  
CC ENDOPLASMIC RETICULUM AND THE GOLGI APPARATUS.  
CC -1- TISSUE SPECIFICITY: BRAIN AND PARIENTAL CELLS.  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
CC  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC and email to: [license@isb-sib.ch](mailto:license@isb-sib.ch)).

of send all email to: [cc@cc.ccr.ac.uk](mailto:cc@cc.ccr.ac.uk)

|    |                                                           |
|----|-----------------------------------------------------------|
| CC | EMBL; X68071; CAA48208.1; -.                              |
| CC | EMBL; X68071; CAA48208.1; -.                              |
| DR | PIR; S23979; S23979.                                      |
| DR | PIR; S23979; S23979.                                      |
| DR | HSP; P05713; 3RAB.                                        |
| DR | HSP; P05713; 3RAB.                                        |
| DR | InterPro; IPR003579; GTPase_Rab.                          |
| DR | InterPro; IPR003579; GTPase_Rab.                          |
| DR | InterPro; IPR001806; Ras_trnsmng.                         |
| DR | InterPro; IPR001806; Ras_trnsmng.                         |
| DR | InterPro; IPR005225; Small_GTP.                           |
| DR | InterPro; IPR005225; Small_GTP.                           |
| DR | Pfam; PF00071; ras; 1.                                    |
| DR | Pfam; PF00071; ras; 1.                                    |
| DR | PRINTS; PR00449; RASTNSFRMNG.                             |
| DR | PRINTS; PR00449; RASTNSFRMNG.                             |
| DR | SMART; SM00175; RAB; 1.                                   |
| DR | SMART; SM00175; RAB; 1.                                   |
| DR | TIGRFS; TIGR00231; small_gtp; 1.                          |
| DR | TIGRFS; TIGR00231; small_gtp; 1.                          |
| KW | GTP-binding; Lipoprotein; Prenylation; Protein transport. |
| KW | GTP-binding; Lipoprotein; Prenylation; Protein transport. |
| FT | NP_BIND 13 20 GTP (BY SIMILARITY).                        |
| FT | NP_BIND 13 20 GTP (BY SIMILARITY).                        |
| FT | NP_BIND 61 65 GTP (BY SIMILARITY).                        |
| FT | NP_BIND 61 65 GTP (BY SIMILARITY).                        |

100

1. *Phragmites australis* (Cav.) Trin. ex Steud.

FT NP\_BIND 119 122 GTP (BY SIMILARITY).  
FT DOMAIN 35 43 EFFECTOR REGION (BY SIMILARITY).  
FT LIPID 211 212 GERANYL-GERANYL (BY SIMILARITY).  
FT LIPID 212 212 GERANYL-GERANYL (BY SIMILARITY).  
SQ SEQUENCE 212 AA; 23531 MW; F8731E3F8B689A3 CRC64;  
Query Match 42.1%; Score 539; DB 1; Length 212;  
Best Local Similarity 56.0%; Pred. No. 1.5e-38;  
Matches 103; Conservative 30; Mismatches 51; Indels 0; Gaps 0;  
QY 42 FLKELVIGSAGCKSCLLHQFTENKFKQDSNHTIGVEFSRVVNVGKVKLQIWDTAG 101  
DB 5 YLFKIIIGTGVKSCLLQFTDKRFQPVHDLTIGVEFGARMITIDGKQIKLQIWDTAG 64  
QY 102 QERFSRVTSYRGAAGALLVYDITSRETYSNLAALWTDARTLASPNVIVLGCNKKDLD 161  
DB 65 QESFRSITRYSYRGAAGALLVYDITRDITFNHLTWLEDARQHSNSNMVIMLGNKSDLE 124  
QY 162 PEREVTFLASRAFAENELMFLTSALTGENVEAFKLCARTILNKIDSGELDPERMGS 221  
DB 125 SREVKKEGEAREHGLIFMETSAKTASNVAEAFINTAKEIYKIQGVDFINNEANG 184  
QY 222 IQYG 225  
DB 185 IKIG 188  
RESULT 14  
RAB2\_LYMST STANDARD; PRT; 212 AA.  
AC Q05975;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE Ras-related protein Rab-2.  
GN RAB2.  
OS Lymnaea stagnalis (Great pond snail).  
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
OC Lymnaeidae; Lymnaea.  
OX NCBI\_TaxID=6523;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE-94039042; PubMed-8223561; Andriessen J.A.;  
RA Agterberg M., van Die I., Yang H.;  
RA van Tetering A., van den Eijnden D.H., Ploegh H.L.;  
RT "Isolation and characterization of three cDNAs coding for Rab  
proteins from the albumen gland of the mollusc Lymnaea stagnalis.";  
RL Eur. J. Biochem. 217:241-246(1993).  
CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
TRAFFIC (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH A STRUCTURE HAVING THE  
CHARACTERISTICS OF AN INTERMEDIATE COMPARTMENT BETWEEN THE  
ENDOPLASMIC RETICULUM AND THE GOLGI APPARATUS (BY SIMILARITY).  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X72689; CAA51234.1;  
DR PIR; S32208; S32208.  
DR PIR; S38341; S38341.  
DR HSSP; P05713; 3RAB.  
DR InterPro; IPR003579; GTPase\_Rab.  
DR InterPro; IPR001806; Ras\_trnsfrmg.  
DR Pfam; PF00071; ras; 1.  
DR PRINTS; PR00449; RASTRNSFRMG.  
DR SMART; SM00175; RAB; 1.

TIGRAMS; TIGR00231; small\_GTP; 1.  
KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
FT NP\_BIND 13 20 GTP (BY SIMILARITY).  
FT NP\_BIND 61 65 GTP (BY SIMILARITY).  
FT NP\_BIND 119 122 GTP (BY SIMILARITY).  
FT DOMAIN 35 43 EFFECTOR REGION (BY SIMILARITY).  
FT LIPID 211 212 GERANYL-GERANYL (BY SIMILARITY).  
FT LIPID 212 212 GERANYL-GERANYL (BY SIMILARITY).  
SQ SEQUENCE 212 AA; 23534 MW; F5878EDBA976752C CRC64;  
Query Match 42.0%; Score 538; DB 1; Length 212;  
Best Local Similarity 56.0%; Pred. No. 1.8e-38;  
Matches 103; Conservative 30; Mismatches 51; Indels 0; Gaps 0;  
QY 42 FLKELVIGSAGCKSCLLHQFTENKFKQDSNHTIGVEFSRVVNVGKVKLQIWDTAG 101  
DB 5 YLFKIIIGTGVKSCLLQFTDKRFQPVHDLTIGVEFGARMITIDGKQIKLQIWDTAG 64  
QY 102 QERFSRVTSYRGAAGALLVYDITSRETYSNLAALWTDARTLASPNVIVLGCNKKDLD 161  
DB 65 QESFRSITRYSYRGAAGALLVYDITRDITFNHLTWLEDARQHSNSNMVIMLGNKSDLE 124  
QY 162 PEREVTFLASRAFAENELMFLTSALTGENVEAFKLCARTILNKIDSGELDPERMGS 221  
DB 125 ARREVKKEGEAREHGLIFMETSAKTASNVAEAFINTAKEIYKIQGVDFINNEANG 184  
QY 222 IQYG 225  
DB 185 IKIG 188  
RESULT 15  
RAB2\_RAT STANDARD; PRT; 212 AA.  
AC P05712;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE Ras-related protein Rab-2.  
GN RAB2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE-88068563; PubMed-3317403;  
RA Touchot N., Chardin P., Tavittian A.;  
RT "Four additional members of the ras gene superfamily isolated by an  
oligonucleotide strategy: molecular cloning of ypt-related cDNAs from  
a rat brain library.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:8210-8214(1987).  
CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
TRAFFIC (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH A STRUCTURE HAVING THE  
CHARACTERISTICS OF AN INTERMEDIATE COMPARTMENT BETWEEN THE  
ENDOPLASMIC RETICULUM AND THE GOLGI APPARATUS (BY SIMILARITY).  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; J02999; AAA42007.1;  
DR PIR; B39963; B39963.  
DR HSSP; P05713; 3RAB.  
DR InterPro; IPR003579; GTPase\_Rab.  
DR InterPro; IPR001806; Ras\_trnsfrmg.  
DR PRINTS; PR00449; RASTRNSFRMG.  
DR SMART; SM00175; RAB; 1.

DR Pfam; PF00071; ras; 1.  
DR PRINTS; PRO0449; RASTRNSFRMG.  
DR SMART; SM00175; RAB; 1.  
DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
FT NP\_BIND 13 20 GTP (BY SIMILARITY).  
FT NP\_BIND 61 65 GTP (BY SIMILARITY).  
FT NP\_BIND 119 122 GTP (BY SIMILARITY).  
FT DOMAIN 35 43 EFFECTOR REGION (BY SIMILARITY).  
FT LIPID 211 211 GERANYL-GERANYL (BY SIMILARITY).  
FT LIPID 212 212 GERANYL-GERANYL (BY SIMILARITY).  
SQ SEQUENCE 212 aa; 23535 MW; B68AAFF894DBE819 CRC64;  
Query Match 41.9%; Score 536; DB 1; Length 212;  
Best Local Similarity 55.4%; Pred. No. 2.7e-38;  
Matches 102; Conservative 31; Mismatches 51; Indels 0; Gaps 0;  
QY 42 FLFKFLVIGSAGTGKSCLLHQFIENKFKODSNHTIGVEFGSRVNVVGGKTVKQLQIWDTAG 101  
DB 5 YLFKIIIGDTGVGKSCLLQLQFTDKRFQPVHDLTWGVEFGARMITDQKQIKLQIWDTAG 64  
QY 102 QERFRSVTRSYRGAAGALLVYDITSRETYNSLAAWLTDARTLASPNVIVILCGNKKDL 161  
DB 65 QESFRSITRSYRGAAGALLVYDITRDTFENHLLTWLEDARQHSNSNMVIMLIGNKSDLE 124  
QY 162 PEREVTFLASRFAQENELMFLETSAITGENVEEAFKLCARTILANKIDSGELDPERMMSG 221  
DB 125 SRREVKKEGEAFAREHGLIFMETSAKTASNVEEAFINTAKEIYEKIQGVFDINNEANG 184  
QY 222 IQYG 225  
DB 185 IKIG 188

Search completed: March 4, 2003, 15:00:03  
Job time: 6.93644 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model.

Run on: March 4, 2003, 14:53:32 : Search time 16.9905 Seconds  
(without alignments)  
3007.543 Million cell updates/sec

Title: US-09-763-902B-8  
Perfect score: 1280  
Sequence: 1 MSVSLPLTWVRERDWIGH.....LRQLRPSAQAVAPQCGC 248

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp Vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 1280  | 100.0       | 248    | 4 Q9HBR6  | Q9hbr6 homo sapien |
| 2          | 1045  | 81.6        | 213    | 11 Q912K1 | Q912k1 mus musculu |
| 3          | 947   | 74.0        | 218    | 4 Q9BQ44  | Q9bq44 homo sapien |
| 4          | 882.5 | 68.9        | 198    | 4 Q9H0Z8  | Q9h0z8 homo sapien |
| 5          | 826   | 64.7        | 213    | 5 Q9V3L5  | Q9v3l5 drosophila  |
| 6          | 738   | 57.7        | 213    | 5 Q44213  | O44213 trichinella |
| 7          | 645.5 | 50.4        | 215    | 11 Q91V41 | Q91v41 mus musculu |
| 8          | 622.5 | 48.6        | 215    | 5 Q18336  | Q18336 drosophila  |
| 9          | 617.5 | 48.2        | 215    | 5 Q8T046  | Q8t046 drosophila  |
| 10         | 606   | 47.3        | 210    | 5 Q93874  | Q93874 caenorhabdi |
| 11         | 586.5 | 45.8        | 223    | 5 Q9GP33  | Q9gp33 echinococcu |
| 12         | 550.5 | 43.0        | 213    | 5 Q9U5D5  | Q9u5d5 drosophila  |
| 13         | 550   | 43.0        | 211    | 10 Q38922 | Q38922 arabidopsis |
| 14         | 546   | 42.7        | 216    | 4 Q8WUD1  | Q8wud1 homo sapien |
| 15         | 545.5 | 42.6        | 213    | 5 Q18333  | Q18333 drosophila  |
| 16         | 544   | 42.5        | 210    | 10 Q50019 | Q50019 zea mays (m |

|    |       |      |     |           |                     |
|----|-------|------|-----|-----------|---------------------|
| 17 | 544   | 42.5 | 211 | 10 Q9SLP2 | Q9slp2 carica papa. |
| 18 | 543   | 42.4 | 210 | 10 Q9XEE1 | Q9xee1 sporobolus   |
| 19 | 540.5 | 42.2 | 216 | 11 Q9DB48 | Q9db48 mus musculu  |
| 20 | 539   | 42.1 | 212 | 13 Q90965 | Q90965 gallus gall  |
| 21 | 539   | 42.1 | 214 | 5 Q01577  | Q01577 caenorhabdi  |
| 22 | 537   | 42.0 | 223 | 10 Q43596 | Q43596 oryza sativ  |
| 23 | 536   | 41.9 | 211 | 10 Q946G3 | Q946g3 nicotiana t  |
| 24 | 535.5 | 41.8 | 213 | 5 Q9BHT6  | Q9bht6 plasmodium   |
| 25 | 534   | 41.7 | 211 | 10 Q40208 | Q40208 lotus japon  |
| 26 | 534   | 41.7 | 211 | 10 P92963 | P92963 arabidopsis  |
| 27 | 527   | 41.2 | 211 | 10 Q39824 | Q39824 glycine max  |
| 28 | 520   | 40.6 | 214 | 5 Q9G078  | Q9g078 giardia lam  |
| 29 | 519.5 | 40.6 | 211 | 5 Q9U9D3  | Q9u9d3 tetrahymena  |
| 30 | 516   | 40.3 | 151 | 11 Q9CYR3 | Q9cyr3 mus musculu  |
| 31 | 504   | 39.4 | 198 | 5 Q26663  | Q26663 trypanosoma  |
| 32 | 503   | 39.3 | 205 | 10 Q23561 | Q23561 arabidopsis  |
| 33 | 489   | 38.2 | 202 | 5 Q8T810  | Q8t810 dictyosteli  |
| 34 | 487   | 38.0 | 207 | 5 Q9U9D2  | Q9u9d2 tetrahymena  |
| 35 | 477.5 | 37.3 | 196 | 10 Q8S6U4 | Q8s6u4 oryza sativ  |
| 36 | 472   | 36.9 | 192 | 5 Q966X2  | Q966x2 giardia lam  |
| 37 | 455   | 35.5 | 211 | 5 Q01803  | Q01803 caenorhabdi  |
| 38 | 453   | 35.4 | 222 | 10 Q39844 | Q39844 glycine max  |
| 39 | 452.5 | 35.4 | 214 | 5 Q18335  | Q18335 drosophila   |
| 40 | 449   | 35.1 | 224 | 10 Q9SMQ6 | Q9smq6 arabidopsis  |
| 41 | 448   | 35.0 | 225 | 10 Q08147 | Q08147 pisum sativ  |
| 42 | 447.5 | 35.0 | 326 | 10 Q9LMR7 | Q9lmr7 arabidopsis  |
| 43 | 447   | 34.9 | 218 | 11 Q9ET14 | Q9et14 rattus norv  |
| 44 | 445   | 34.8 | 242 | 10 Q40192 | Q40192 lotus japon  |
| 45 | 444.5 | 34.7 | 214 | 3 Q8WZR7  | Q8wzr7 neurospora   |

ALIGNMENTS

RESULT 1

Q9HBR6 PRELIMINARY; PRT; 248 AA.  
AC Q9HBR6;  
DT 01-MAR-2001 (TREMREL. 16, Created)  
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)  
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)  
DE Hypothetical 27.5 kDa protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.O., Zhang P.P.,  
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,  
RA Yu J., Han L.H.;  
RT "Novel Human cDNA clones with function of inhibiting cancer cell  
RT growth.";  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
DR EMBL; AF217945; AAG17228.1;  
DR HSSP; P36017; 1EKO.  
DR InterPro; IPR003579; GTPase\_Rab.  
DR InterPro; IPR001806; Ras\_trnsfrim.  
DR InterPro; IPR005225; Small\_GTP.  
DR Pfam; PF00071; ras; 1.  
DR PRINTS; PR00449; RASTRNSFRMG.  
DR SMART; SM00175; RAB; 1.  
DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
KW GTP-binding; Hypothetical protein; Lipoprotein.  
SQ SEQUENCE 248 AA; 27506 MW; 69EBDC77A684EAC6 CRC64;

Query Match 100.0%; Score 1280; DB 4; Length 248;  
Best Local Similarity 100.0%; Pred. No. 1.8e-108;  
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MSVSLPLTWVRERDWIGHLFSVLSPVGPDPGSIWSDFLFKFLVIGSAGTKSCLL 60  
|||||

```

Db 1 MSVSLPLTVWVRDWDIGIHLSFLSLVLPVGPDPFGSIHSDFLFKFLVIGSAGTKSCILL 60
QY 61 HOFIENKFKODSNHTTIGVFGSRVNVGKTKVLQIWDTAGOERFRSVTSYRGAAGAL 120
Db 61 HOFIENKFKODSNHTTIGVFGSRVNVGKTKVLQIWDTAGOERFRSVTSYRGAAGAL 120
QY 121 LVYDITSRETYNSLAWLTDARTLASPNIVILCGNKKDLDPEREVTFLEASRFAOENEL 180
Db 121 LVYDITSRETYNSLAWLTDARTLASPNIVILCGNKKDLDPEREVTFLEASRFAOENEL 180
QY 181 MFLTSALTGENVEEAFKACARTILNKIDSGELDPERMGSQIYGDASRLQRLQRPASQA 240
Db 181 MFLTSALTGENVEEAFKACARTILNKIDSGELDPERMGSQIYGDASRLQRLQRPASQA 240
QY 241 VAPQPCGC 248
Db 241 VAPQPCGC 248

RESULT 2
Q91ZR1 PRELIMINARY; PRT; 213 AA.
AC Q91ZR1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE GTP-binding protein RAB4.
GN RAB4B OR RAB4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhou M., Raschke W.C.;
RT "Mus musculus GTP-binding protein (RAB4) mRNA.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF048432; AAL11725.1;
DR MGD; MGI:105071; Rab4b.
DR InterPro; IPR001806; Ras_trnsfrmng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
KW GTP-binding.
SQ SEQUENCE 213 AA; 23554 MW; 9FF249DC329E0A27 CRC64;

Query Match 81.6%; Score 1045; DB 11; Length 213;
Best Local Similarity 98.1%; Pred. No. 3.6e-87;
Matches 204; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 41 DFLKFLVIGSAGTKSCILLHOFIENKFKODSNHTTIGVFGSRVNVGKTKVLQIWDTA 100
Db 6 DFLKFLVIGSAGTKSCILLHOFIENKFKODSNHTTIGVFGSRVNVGKTKVLQIWDTA 65
QY 101 GOERFSVTSYRGAAGALLVYDITSRETYNSLAWLTDARTLASPNIVILCGNKKDL 160
Db 66 GOERFSVTSYRGAAGALLVYDITSRETYNSLAWLTDARTLASPNIVILCGNKKDL 125
QY 161 DPEREVTFLEASRFAOENELMFLTSALTGENVEEAFKACARTILNKIDSGELDPERMGS 220
Db 126 DPEREVTFLEASRFAOENELMFLTSALTGENVEEAFKACARTILNKIDSGELDPERMGS 185
QY 221 GYOYGDASRLQRLQRPASQAQVAPQPCGC 248
Db 186 GYOYGDISRLQRLHARSQAQVAPQPCGC 213

RESULT 3
Q9BQ44 PRELIMINARY; PRT; 218 AA.
AC Q9BQ44;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

```

```

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE RAB4, member RAS oncogene family.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; BC004309; AAH04309.1;
DR EMBL; BC002438; AAH02438.1;
DR HSSP; P36017; IEK0.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsfrmng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PRO0449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 218 AA; 24389 MW; 983D65E1162741B0 CRC64;

Query Match 74.0%; Score 947; DB 4; Length 218;
Best Local Similarity 87.5%; Pred. No. 3.1e-78;
Matches 182; Conservative 11; Mismatches 15; Indels 0; Gaps 0;

QY 41 DFLKFLVIGSAGTKSCILLHOFIENKFKODSNHTTIGVFGSRVNVGKTKVLQIWDTA 100
Db 11 DFLKFLVIGSAGTKSCILLHOFIENKFKODSNHTTIGVFGSRVNVGKTKVLQIWDTA 70
QY 101 GOERFSVTSYRGAAGALLVYDITSRETYNSLAWLTDARTLASPNIVILCGNKKDL 160
Db 71 GOERFSVTSYRGAAGALLVYDITSRETYNSLAWLTDARTLASPNIVILCGNKKDL 130
QY 161 DPEREVTFLEASRFAOENELMFLTSALTGENVEEAFKACARTILNKIDSGELDPERMGS 220
Db 131 DADREVTTFLEASRFAOENELMFLTSALTGENVEEAFVOCARKILNKIESGELDPERMGS 190
QY 221 GYOYGDASRLQRLQRPASQAQVAPQPCGC 248
Db 191 GYOYGDALRLQRLSPRAQAPNAQECGC 218

RESULT 4
Q9H028 PRELIMINARY; PRT; 198 AA.
AC Q9H028;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE DJ803J11.1 (RAB4, member RAS oncogene family) (Fragment).
GN RAB4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hall R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; AL117350; CAC17572.1;
DR HSSP; P05713; 12BD.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsfrmng.

```

```

DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; ras; 1.
DR PRINTS: PRO0449; RASTRNSFRMNG.
DR SMART: SM00175; RAB; 1.
DR TIGRFAMs: TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein.
FT NON_TER 1
SQ SEQUENCE 198 AA; 22037 MW; 566A2C25BB57A496 CRC64;

Query Match 68.9%; Score 882.5; DB 4; Length 198;
Best Local Similarity 84.7%; Pred. No. 28-72;
Matches 172; Conservative 11; Mismatches 15; Indels 5; Gaps 1;

QY 46 FLVIGSAGTGKSCLLHOFIENFKODSNHTIGVEFGSRVNVGVKQVQIWDTAQGERF 105
DB 1 FLVIGNAGTGKSCLLHOFIEKFKDSDNHTIGVEFGSKINNVGVKQVQIWDTAQGERF 60

QY 106 RSVTRSYRGAAGALVYDITSRETYSNLAAMLTADARTLASPNVIVILCGNKKDLDPERE 165
DB 1 RSVTRSYRGAAGALVYDITSRETYSNLAAMLTADARTLASPNVIVILCGNKKDLDPERE 165

QY 61 RSVTRSYRGAAGALVYDITSRETYSNLAAMLTADARTLASPNVIVILCGNKKDLDPERE 120
DB 1 RSVTRSYRGAAGALVYDITSRETYSNLAAMLTADARTLASPNVIVILCGNKKDLDPERE 120

QY 166 VTFLEASRAQENELMFLTSALTGENVEAFKLCARTILNKIDSGELDPERMGSQIYQ 225
DB 121 VTFLEASR-----LMFLTSALTGENVEAFVQCARKILNKIESGELDPERMGSQIYQ 175

QY 226 DASLRLQRPRAQAQVAPQPCG 248
DB 176 DAALRLSPRAQAQVAPQPCG 198

RESULT 5
QYV3L5 PRELIMINARY; PRT; 213 AA.
AC QYV3L5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE RAB4 protein.
GN RAB4 OR CG4921.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

```

```

RL Parasitol. Res. 85:607-611(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-GARKAVI,
RA Garate T.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC EMBL: Y1181; CAA72626.2;
DR HSSP: P05713; 3RAB.
DR InterPro: IPR003579; GTPase_Rab.
DR InterPro: IPR001806; Ras_trnsmg.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; ras; 1.
DR PRINTS: PR00449; RASTNSFRNG.
DR SMART: SM00175; RAB; 1.
DR TIGRFAMS: TIGR00231; small_gtp; 1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 213 AA; 24065 MW; 48BA45507605756D CRC64;

Query Match 57.7%; Score 738; DB 5; Length 213;
Best Local Similarity 67.2%; Pred. No. 3.2e-59;
Matches 131; Conservative 41; Mismatches 23; Indels 0; Gaps 0;

QY 42 FLFKLVIGSAGTGGKSCLLHQFTENFKQDSNHTIGVEGSRVYVGGTKVQLQIWDTAG 101
DB 7 YLKYLIIGNSGTGKSCVLLHQFTENFKQDTAHTIGLEFGSKVVKIQDRSIKQLQVWDTAG 66

QY 102 QERFRSVTRYRGAAGALVYDITSRETYNSLAAMTDTARTLASPNVILCGNKKDL 161
DB 67 QERFSUTKRYNGACALVYDITCRQSNATQWLSARSQIIVILGNKKDL 126

QY 162 PEREVTLEASRAFAQENLMFLTSALTGENVEAFKLCARTILNKIDSGELDPERM 221
DB 127 DRREVTMEASQFAQENGLFLETSALTGENIEBTEFLRCARSILTKIESGELDPN 186

QY 222 IOYGDASLRQLRROP 236
DB 187 IQGHVEMKPIKND 201

RESULT 7
Q91V41 PRELIMINARY; PRT; 215 AA.
AC Q91V41: 2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Adult male kidney cDNA, RIKEN full-length enriched library,
DE clone:0610030G24, full insert sequence (Unknown) (Protein for
DE MGC:6512).
DE Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-KIDNEY;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Aono H., Arai A.,
RA Arakawa T., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kato H., Kawai J., Kojima Y.,
RA Konno H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Okazaki Y., Okido T.,
RA Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D.,
RA Shibata K., Shibata Y., Shingawa A., Shiraki T., Sogabe Y.,
RA Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y.,
RA Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-KIDNEY;

```

```

RX MEDLINE-21085660; PubMed-11217851;
RA RIKEN FANTOM Consortium.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-KIDNEY;
RX MEDLINE-99279253; PubMed-10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-KIDNEY;
RX MEDLINE-20499374; PubMed-11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-KIDNEY;
RX MEDLINE-20530913; PubMed-11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kusunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE-BREAST TUMOR;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK002704; BAB22298.1;
DR EMBL: BC009085; AAH09085.1;
DR MGI: 1915615; Rab14.
DR InterPro: IPR001806; Ras_trnsmg.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; ras; 1.
DR TIGRFAMS: TIGR00231; small_gtp; 1.
KW GTP-binding.
SQ SEQUENCE 215 AA; 23897 MW; BC8A8B98FB9944AC CRC64;

Query Match 50.4%; Score 645.5; DB 11; Length 215;
Best Local Similarity 58.0%; Pred. No. 8.5e-51;
Matches 123; Conservative 33; Mismatches 45; Indels 11; Gaps 3;

QY 42 FLFKLVIGSAGTGGKSCLLHQFTENFKQDSNHTIGVEGSRVYVGGTKVQLQIWDTAG 101
DB 10 YIFKIIIGDMGVGKSCLLHQFTENFKQDPCPHITGVEGTRIEVSGQIKQLQIWDTAG 69

QY 102 QERFRSVTRYRGAAGALVYDITSRETYNSLAAMTDTARTLASPNVILCGNKKDL 161
DB 70 QERFVTRYRGAAGALVYDITRETYNHLSSWLTARNLTNPNTVILGNKADLE 129

QY 162 PEREVTLEASRAFAQENLMFLTSALTGENVEAFKLCARTILNKIDSGELDPERM 221
DB 130 AQRDVTEEAQFAEENGFLLEASAKTGENVEDAFLEAAKTIYQNTQDGSLDLNAE 189

QY 222 IOYGDASLRQLRROP 236
DB 190 VQHRPSA-----PQGRLLISEPQREGGCC 215

RESULT 8
Q18336 PRELIMINARY; PRT; 215 AA.
ID Q18336

```

AC 018336;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE RAB14 protein.  
 GN RAB14 OR C04212.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Branton R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harkin N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattle B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.E.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=99403001; PubMed=10471707;  
 RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,  
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,  
 RA Hong L., Houston K., Hoskins R., Johnson C., Martin C., Moshrefi A.,  
 RA Palazolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,  
 RA Celniker S., Rubin G.M.;  
 RT "An exploitation of the sequence of a 2.9-Mb region of the genome of  
 Drosophila melanogaster: the Adh region."  
 RL Genetics 153:179-219(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,  
 RA Butenoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,  
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.,  
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,  
 RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,

RA Nixon K., Pacle J.E., Park S., Pfeiffer B., Poon L., Sequeira A.,  
 RA Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,  
 RA Zierán L.L., Rubin G.M.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OREGON R; TISSUE-HEAD;  
 RX MEDLINE=97228579; PubMed=9074639;  
 RA Satch A.K., Tokunaga F., Ozaki K.;  
 RT "Rab proteins of Drosophila melanogaster: novel members of the Rab-  
 protein family."  
 RL FEBS Lett. 404:65-69(1997).  
 CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL: AE003643; AAF53390.1;  
 DR EMBL: AE003409; AAF44870.1;  
 DR EMBL: D84316; BAA21709.1;  
 DR HSSP: P05713; 3RAB.  
 DR FLYBase: FBgn0015791; Rab14.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnstrmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRMNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAMs: TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein.  
 SQ SEQUENCE 215 AA; 24256 MW; 88A09B7FFB41637F CRC64;  
 Query Match 48.6%; Score 622.5; DB 5; Length 215;  
 Best Local Similarity 57.9%; Pred. No. 1.1e-48;  
 Matches 121; Conservative 29; Mismatches 56; Indels 3; Gaps 2;  
 QY 41 DFLFELVIGSAGTCKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGKTVKQLWDTA 100  
 DB 9 NYIFRYIIIGDMGVGKSCLLHOFTEKFKMNCPTHTIGVEFGTRIEVDKIKLQWDTA 68  
 QY 101 GOERFSRVSRYRGAAGALLVDITSTRTYNSLAWLTDARTLASPNIVTILCGNKKDL 160  
 DB 69 GOERFVTRSYRGAAGALMVDITRSTYNSLWLTDRNLNTPSVIFLGNKSDL 128  
 QY 161 DPERVTFLEASRFAQENELMLETSLTNGENVEFAFLKCAITLKNKIDSGELDPERMS 220  
 DB 129 ESTREVTVEAEKEFADENGLMFEASAMTGVNVEAFLEATARKIYQIQEGRDLNASES 188  
 QY 221 GTOYGDASLRQLRQPSRAQV-APQPCG 248  
 DB 189 GVQHRPS--QPSRSLSEATGAKDQCSC 215  
 RESULT 9  
 QST046  
 ID Q8T046 PRELIMINARY; PRT; 215 AA.  
 AC Q8T046;  
 DT 01-JUN-2002 (TRENBLrel. 21, Created)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE LD29476P.  
 GN RAB14.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuncio J., Pacle J.E., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 EMBL: AY069563; AAL39708.1;



```

SQ SEQUENCE 215 AA; 24230 MW; 9EB59B7FEE546CDA CRC64;
Query Match 48.2%; Score 617.5; DB 5; Length 215;
Best Local Similarity 57.4%; Pred. No. 3e-48;
Matches 120; Conservative 30; Mismatches 56; Indels 3; Gaps 2;

QY 41 DFLFKFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGGTKVKLQIWDTA 100
Db 9 NVIFKVIIGDMGVGKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGGTKVKLQIWDTA 68

QY 101 GQERFSVTRSYRGAAGALLVDITSRITYNSLAALWLDARTLASPNVIVILCGNKKDL 160
Db 69 GQERFSVTRSYRGAAGALLVDITSRITYNSLAALWLDARTLASPNVIVILCGNKKDL 128

QY 161 DPERVTFLEASFAOENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 220
Db 129 ESTRETYEAEKFADENGIMLEASAMTGVNVEEAFLETKRIYQNIQEGRLDNLASES 188

QY 221 GYOYDASLRQLRQPSAQAV-APQPCGC 248
Db 189 GVQHRPS--QPSRTSLSEATGAKQDCSC 215

RESULT 10
Q93874
ID Q93874 PRELIMINARY; PRT; 210 AA.
AC Q93874;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE K09A9.2 protein.
GN K09A9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Swinburne J.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
SQ SEQUENCE FROM N.A.
MEDLINE=99089613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; Z79601; CAB01884.1;
DR HSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; Ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_gtp; 1.
DR GTP-binding; Lipoprotein.
KW SEQUENCE 210 AA; 23388 MW; DDE4B52E447656D CRC64;

Query Match 47.3%; Score 606; DB 5; Length 210;
Best Local Similarity 55.1%; Pred. No. 3.3e-47;
Matches 114; Conservative 32; Mismatches 55; Indels 6; Gaps 1;

QY 42 DFLFKFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGGTKVKLQIWDTA 101
Db 10 YIFKVIIGDMGVGKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGGTKVKLQIWDTA 69

QY 102 QERFSVTRSYRGAAGALLVDITSRITYNSLAALWLDARTLASPNVIVILCGNKKDL 161
Db 70 QERFSVTRSYRGAAGALLVDITSRITYNSLAALWLDARTLASPNVIVILCGNKKDL 129

QY 162 PERVTFLEASFAOENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221
```

```

Db 130 DORDVPYEEAKAFABENGLTFLECSAKTGSNVDAFLATAKIYQNIQDGSLLDAADTG 189
QY 222 IOYGDASLRQLRQPSAQAVAPQPCGC 248
Db 190 VQ-----PKQNLPRAAENNGKKKNC 210

RESULT 11
Q9GP33
ID Q9GP33 PRELIMINARY; PRT; 223 AA.
AC Q9GP33;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative ras-related protein RAB4A.
GN RAB4A.
OS Echinococcus multilocularis.
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Echinococcus.
OX NCBI_TaxID=6211;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-H-95;
RX MEDLINE=20556267; PubMed=10973970;
RA Brehm K.; Jensen K.; Frosch M.;
RT "mRNA trans-splicing in the human parasitic cestode Echinococcus
multilocularis."
RL J. Biol. Chem. 275:38311-38318(2000).
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; AJ292375; CAC18549.1;
DR HSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; Ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_gtp; 1.
DR GTP-binding; Lipoprotein.
KW SEQUENCE 223 AA; 24454 MW; 9EB6D6166E2089A4 CRC64;

Query Match 45.8%; Score 586.5; DB 5; Length 223;
Best Local Similarity 56.6%; Pred. No. 2.1e-45;
Matches 116; Conservative 37; Mismatches 45; Indels 7; Gaps 3;

QY 41 DFLFKFLVIGSAGTGKSCLLHOFIENKFKQDSNHTIGVEFGSRVNVGGTKVKLQIWDTA 100
Db 7 DFLFKFLIIGNAGTGKTCILRRYTERKFFPTQHTIGAEGFSRVISVDGTHVKIWDTA 66

QY 101 GQERFSVTRSYRGAAGALLVDITSRITYNSLAALWLDARTLASPNVIVILCGNKKDL 160
Db 67 GQERFSVTRSYRGAAGALLVDITSRITYNSLAALWLDARTLASPNVIVILCGNKKDL 126

QY 161 DPERVTFLEASFAOENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERM- 218
Db 127 RDTGQVTHWEANTFAENGLOFIETSLTGENIDDAFTSCVRVLLSKVSGELGADRL 186

QY 219 -GSGIQYGDASLRQLRQPSAQAV 242
Db 187 VGSNKQH----LQAVNLTSATSVS 207

RESULT 12
Q9U5D5
ID Q9U5D5 PRELIMINARY; PRT; 213 AA.
AC Q9U5D5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE DRAB2.
GN RAB2 OR DRAB2 OR CG3269.
OS Drosophila melanogaster (Fruit fly).
```

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sasamura T.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97340933; PubMed=9197407;  
 RA Sasamura T., Kobayashi T., Kojima S., Qadota H., Ohya Y., Masai I.,  
 RA Hotta Y.;  
 RT "Molecular cloning and characterization of Drosophila genes encoding  
 small GTPases of the rab and rho families.";  
 RL Mol. Gen. Genet. 254:486-494(1997).  
 CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL; AB035352; BAB87878.1;  
 DR HSP; P05713; 3RAB.  
 DR FlyBase; FBgn0014009; Rab2.  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsmfmg.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTRNSFRMG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRfams; TIGR00231; small\_GTP; 1.  
 DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 KW GTP-binding; Lipoprotein.  
 SQ SEQUENCE 213 AA; 23537 MW; 669B03897679CC02 CRC64;  
 Query Match 43.0%; Score 550.5; DB 5; Length 213;  
 Best Local Similarity 53.4%; Pred. No. 3.8e-42;  
 Matches 111; Conservative 28; Mismatches 68; Indels 1; Gaps 1;  
 OY 42 FLFLKLVTSAGTSGKSCLLHQFIENKFKQDSNHTTIGVFGSRVNVGGTKVQLQIWDTAG 101  
 DB 5 YLFKVIIGDTGVGKSCLLQFTDRFQPVHDLTIGVEFGARMITDGRQIKLQIWDTAG 64  
 OY 102 QERFVSRTSYRGAAGALLVYDITSRTYNSLAALWLTDAITLASPNVIVLCGNKKDL 161  
 DB 65 QEAFTSRTSYRGAAGALLVYDITRRTFTNLTWLEDARQHSNSNMVIMLIGNKSDLD 124  
 OY 162 PREVTFLFASFAQENELMFLTSAITGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
 DB 125 SRREYKKEGEAFRENGLVFMETSAITAAVVEEAFINTAKEIYEKIQGVDFDINNEANG 184  
 OY 222 IQYGDG-SLRQLRQPSAQAVAPQCGC 248  
 DB 185 IKIGQHSPTNPSLPGAGGAAGANSGC 212  
 RESULT 13  
 Q38922 PRELIMINARY; PRT; 211 AA.  
 AC Q38922;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-JUN-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE GTP-binding protein ATG82 (GTP-binding protein GB2) (Putative)  
 DE GTP-binding protein GB2.  
 GN F4B14.130 OR AT4G35860.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Biermann B.J., Randall S.K., Crowell D.N.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA Bevan M., Rose M., Hempel S., Entian K.-D., Hohelsel J., Jesse T.,  
 RA Heijnen L., Vos P., Meves H.W., Mayer K.F.X., Schueller C.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Rose M., Hempel S., Entian K.-D., Meves H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,  
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,  
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
 RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Full length cDNA of gene At4g35860 (GI:15233367).";  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL; U46925; AAA87883.1;  
 DR EMBL; AL031986; CAA21472.1;  
 DR EMBL; AL161588; CAB81495.1;  
 DR EMBL; AY065262; AAL38738.1;  
 DR HSP; P05713; 3RAB.  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsmfmg.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTRNSFRMG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRfams; TIGR00231; small\_GTP; 1.  
 DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 KW GTP-binding; Lipoprotein.  
 SQ SEQUENCE 211 AA; 23175 MW; 71B5036D4B9A158E CRC64;  
 Query Match 43.0%; Score 550; DB 10; Length 211;  
 Best Local Similarity 58.4%; Pred. No. 4.1e-42;  
 Matches 108; Conservative 29; Mismatches 48; Indels 0; Gaps 0;  
 OY 41 DFLFPLVTSAGTSGKSCLLHQFIENKFKQDSNHTTIGVFGSRVNVGGTKVQLQIWDTA 100  
 DB 4 DYLFKVIIGDTGVGKSCLLQFTDRFQPVHDLTIGVEFGARMITDGRQIKLQIWDTA 63  
 OY 101 GOERFSRTSYRGAAGALLVYDITSRTYNSLAALWLTDAITLASPNVIVLCGNKKDL 160  
 DB 64 GOESFRSRTSYRGAAGALLVYDITRRTFTNLTWLEDARQHSNSNMVIMLIGNKDL 123  
 OY 161 DPERVTFLFASFAQENELMFLTSAITGENVEEAFKLCARTILNKIDSGELDPERMGS 220  
 DB 124 AKRAVSKEGQGFQFAKEHGLLFLEASARTAQNVVEAFIETAAKILQNTQIDGVDFVDSN 183  
 OY 221 GIQYG 225  
 DB 184 GIKIG 188  
 RESULT 14  
 Q8WUD1 PRELIMINARY; PRT; 216 AA.  
 ID Q8WUD1;  
 AC Q8WUD1;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)

01-JUN-2002 (TReMBLrel\_21, Last annotation update)  
Similar to RIKEN CDNA 1500012D09 gene.  
Homo sapiens (Human).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=SKIN;  
Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
EMBL; BC020839; AAH20839.1;  
InterPro; IPR003579; GIPase\_Rab.  
InterPro; IPR001806; Ras\_trnsfmg.  
InterPrC; IPR002078; Sig54\_interact.  
Pfam; PF00071; ras; 1.  
PRINTS; PR00449; RASTRNSFRMG.  
SMART; SM00175; RAB; 1  
PROSITE; PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1  
SEQUENCE 216 AA; 24214 MW; 80A73F6D259701D8 CRC64;  
  
Query Match 42.7%; Score 546; DB 4; Length 216;  
Best Local Similarity 57.6%; Pred. No. 9.8e-42;  
Matches 106; Conservative 29; Mismatches 49; Indels 0; Gaps  
  
QY 42 FLEKLVIGSAGTGSCLLLHQFIENKFQDSNHTTCGVFGSRVNVGKYKLQIWDTAG 101  
Dbb :|||:||| |||||::: ||: |||||:||||: |||:|||||  
5 YLFKVIIIGDTGVGKSLLLQFTDRKRFVPVHDLTGVEFGARVWIDGKQIKLQIWDTAG 64  
  
QY 102 QERFSRVTSYRGGAALLVVDITSRETYNSLAWLTDARTLASPNIVVILCGNNKKLD 161  
Dbb || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
65 QESFISITSYRGGAALLVVDITRRFTNFHLTWLEDAROHSSNMVIMLIGNKSDLE 124  
  
QY 162 PEREVTEFASFAOENELMFLETSAITCENVEEAFLKCARTILNKIDSGELDPERMGSG 221  
Dbb :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
125 SRDVKREGEAFAREHGLIFMETSKTACNVEEAFFINTAKELYRKIQOGLFDVNEANG 184  
  
QY 222 IQVG 225  
Dbb :|:  
185 IKIG 188  
  
RESULT 15  
O18333 PRELIMINARY; PRT; 213 AA.  
ID O18333 ID O18333 AC O18333  
DT 01-JAN-1998 (TReMBLrel\_05, Created)  
DT 01-JAN-1998 (TReMBLrel\_05, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel\_21, Last annotation update)  
DN RAB2 protein.  
GN RAB2 OR CG3269.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A..  
RC STRAIN=BERKELEY.  
RX MEDLINE=20196006; PubMed=107311132;  
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ball R.M., Basu A., Bakendale J., Bayraktarogl L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mayes A.D., Dew I., Dietz S.M.,



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:55:07 ; Search time 8.59761 Seconds  
(without alignments)  
848.711 Million cell updates/sec

Title: US-09-763-902B-8  
Perfect score: 1280  
Sequence: 1 MSVSLPTVMVRDWMIGH.....LRQLRPSRAQAAPQPCGC 248

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 869.5 | 67.9        | 208    | 2     | US-08-531-525-20  |
| 2          | 869.5 | 67.9        | 208    | 2     | US-08-718-270A-20 |
| 3          | 539   | 42.1        | 212    | 4     | US-09-399-913-67  |
| 4          | 536   | 41.9        | 212    | 2     | US-08-531-525-18  |
| 5          | 536   | 41.9        | 212    | 2     | US-08-718-270A-18 |
| 6          | 508   | 39.7        | 210    | 2     | US-08-531-525-16  |
| 7          | 508   | 39.7        | 210    | 2     | US-08-718-270A-16 |
| 8          | 492.5 | 38.5        | 208    | 2     | US-08-531-525-17  |
| 9          | 492.5 | 38.5        | 208    | 2     | US-08-718-270A-17 |
| 10         | 490   | 38.3        | 208    | 2     | US-08-531-525-15  |
| 11         | 490   | 38.3        | 208    | 2     | US-08-718-270A-15 |
| 12         | 443   | 34.6        | 214    | 2     | US-08-531-525-52  |
| 13         | 443   | 34.6        | 214    | 2     | US-08-718-270A-52 |
| 14         | 425.5 | 33.2        | 213    | 2     | US-08-773-423-8   |
| 15         | 423   | 33.0        | 202    | 2     | US-08-531-525-14  |
| 16         | 423   | 33.0        | 202    | 2     | US-08-718-270A-14 |
| 17         | 421.5 | 32.9        | 217    | 2     | US-08-773-423-3   |
| 18         | 417.5 | 32.6        | 201    | 2     | US-08-916-901-3   |
| 19         | 417.5 | 32.6        | 201    | 4     | US-09-154-602-3   |
| 20         | 417   | 32.6        | 201    | 2     | US-08-916-901-8   |
| 21         | 417   | 32.6        | 201    | 4     | US-09-154-602-8   |
| 22         | 414   | 32.3        | 213    | 2     | US-08-531-525-11  |
| 23         | 414   | 32.3        | 213    | 2     | US-08-718-270A-11 |
| 24         | 410.5 | 32.1        | 201    | 2     | US-08-531-525-13  |
| 25         | 410.5 | 32.1        | 201    | 2     | US-08-718-270A-13 |
| 26         | 408   | 31.9        | 203    | 2     | US-08-766-551-8   |
| 27         | 397.5 | 31.1        | 213    | 2     | US-08-531-525-36  |

|    |       |      |     |   |                   |
|----|-------|------|-----|---|-------------------|
| 28 | 397.5 | 31.1 | 213 | 2 | US-08-718-270A-36 |
| 29 | 393   | 30.7 | 207 | 2 | US-08-824-873-4   |
| 30 | 393   | 30.7 | 207 | 3 | US-09-198-184-4   |
| 31 | 377.5 | 29.5 | 198 | 2 | US-08-531-525-51  |
| 32 | 377.5 | 29.5 | 198 | 2 | US-08-718-270A-51 |
| 33 | 376.5 | 29.4 | 205 | 2 | US-08-531-525-25  |
| 34 | 376.5 | 29.4 | 205 | 2 | US-08-718-270A-25 |
| 35 | 375   | 29.3 | 215 | 2 | US-08-531-525-10  |
| 36 | 375   | 29.3 | 215 | 2 | US-08-718-270A-10 |
| 37 | 373   | 29.1 | 207 | 2 | US-08-531-525-35  |
| 38 | 373   | 29.1 | 207 | 2 | US-08-718-270A-35 |
| 39 | 371.5 | 29.0 | 191 | 4 | US-09-075-454-3   |
| 40 | 369.5 | 28.9 | 218 | 2 | US-08-531-525-19  |
| 41 | 369.5 | 28.9 | 218 | 2 | US-08-718-270A-19 |
| 42 | 365.5 | 28.6 | 190 | 2 | US-08-824-873-3   |
| 43 | 365.5 | 28.6 | 190 | 3 | US-09-198-184-3   |
| 44 | 365   | 28.5 | 194 | 2 | US-08-531-525-34  |
| 45 | 365   | 28.5 | 194 | 2 | US-08-718-270A-34 |

ALIGNMENTS

RESULT 1  
US-08-531-525-20  
; Sequence 20: Application US/08531525  
; Patent No. 5840883  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5840683le, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
; TITLE OF INVENTION: Of P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C. 201  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,525  
; FILING DATE: 21-SEP-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 37-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 208 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Rattus norvegicus  
US-08-531-525-20

Query Match 67.9%; Score 869.5; DB 2; Length 208;  
Best Local Similarity 83.7%; Pred. No. 1.1e-91;

|    | Matches | 174;                                                          | Conservative | 12; | Mismatches | 17; | Indels | 5; | Gaps | 4; |
|----|---------|---------------------------------------------------------------|--------------|-----|------------|-----|--------|----|------|----|
| Qy | 41      | DFLFKFLVIGSAGTGCSCLLHOFIENKFQDSNHTIGVEFGSRVVNVGGTKVKLQIWDTA   | 100          |     |            |     |        |    |      |    |
|    |         | :     :     :     :     :     :     :     :     :     :     : |              |     |            |     |        |    |      |    |
| Db | 6       | DFL-KELVIGNAGTGKSCLLHOFIEKKFKDSDSNHTIGVEFGOKIINVGSKYVKLQIWDTA | 64           |     |            |     |        |    |      |    |
|    |         | :     :     :     :     :     :     :     :     :     :       |              |     |            |     |        |    |      |    |
| Qy | 101     | GOERFRSVTRSYRGAAGALLVYDITSRETYNSLAWLTDARTLASPNIVILCGNKDDL     | 160          |     |            |     |        |    |      |    |
|    |         | :     :     :     :     :     :     :     :     :     :       |              |     |            |     |        |    |      |    |
| Db | 65      | GOERFR-VTTTS-YRGAAGALLVYDITSRETYNALTNWLTDMRLASQNIVI--CGNKDDL  | 120          |     |            |     |        |    |      |    |
|    |         | :     :     :     :     :     :     :     :     :     :       |              |     |            |     |        |    |      |    |
| Qy | 161     | DPEREVTFLASRAEQNELMFLTSLATGENVEAFCLKCARTTLNKTDSELDEPWRMS      | 220          |     |            |     |        |    |      |    |
|    |         | :     :     :     :     :     :     :     :     :     :       |              |     |            |     |        |    |      |    |
| Db | 121     | DADREVTFLEASRAEQNELMFLTSLATGENVEAFMQCCARKILKTESGELODERMGS     | 180          |     |            |     |        |    |      |    |
|    |         | :     :     :     :     :     :     :     :     :     :       |              |     |            |     |        |    |      |    |
| Qy | 221     | GIQYGDAASLRQLRPRAQAQVAPOPCGC                                  | 248          |     |            |     |        |    |      |    |
|    |         | :     :     :     :     :     :     :     :     :     :       |              |     |            |     |        |    |      |    |
| Db | 181     | GIQYGDAALROLRSPRRTOAPSQAQECGC                                 | 208          |     |            |     |        |    |      |    |
|    |         | :     :     :     :     :     :     :     :     :     :       |              |     |            |     |        |    |      |    |

RESULT 2  
HIS-08-718-2702-20

US-08/718-270A-20  
Sequence 20, Application US/08718270A  
Patent No. 5910478  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 59104781e, John F.  
APPLICANT: Abajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptidomimetics Inhibiting  
Title of Invention: The Oncogenic Action of P21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/718,270A  
FILING DATE: 20-SEP-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/531,525  
FILING DATE: 21-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/004,091  
FILING DATE: 21-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 78-95  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 208 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus norvegicus  
US-08/718-270A-20

ORGANISM:  
US-08-718-270A-20

[illegible]

RESULT 3  
US-09-399-913-67  
/ Sequence 67, Application US/09399913  
/ Patent No. 6361971  
/ GENERAL INFORMATION:  
/ APPLICANT: Rhodes, Kenneth  
/ APPLICANT: Betty, Maria  
/ APPLICANT: Ling, Huai-Ping  
/ APPLICANT: An, Wenqian  
/ TITLE OF INVENTION: POTASSIUM CHANNEL INTERACTORS AND USES THEREFOR  
/ FILE REFERENCE: MNI-070CP2  
/ CURRENT APPLICATION NUMBER: US/09/399,913  
/ CURRENT FILING DATE: 1999-09-21  
/ EARLIER APPLICATION NUMBER: USSN 60/110,277  
/ EARLIER FILING DATE: 1998-11-30  
/ EARLIER APPLICATION NUMBER: USSN 60/110,033  
/ EARLIER FILING DATE: 1998-11-25  
/ EARLIER APPLICATION NUMBER: USSN 60/109,333  
/ EARLIER FILING DATE: 1998-11-20  
/ EARLIER APPLICATION NUMBER: USSN 09/298,731  
/ EARLIER FILING DATE: 1999-04-23  
/ EARLIER APPLICATION NUMBER: USSN 09/350,614  
/ EARLIER FILING DATE: 1999-07-09  
/ EARLIER APPLICATION NUMBER: USSN 09/350,874  
/ EARLIER FILING DATE: 1999-07-09  
/ NUMBER OF SEQ ID NOS: 73  
/ SOFTWARE: Patentin Ver. 2.0  
/ SEQ ID NO 67  
/ LENGTH: 212  
/ TYPE: PRT  
/ ORGANISM: Rattus sp.  
US-09-399-913-67

|                       |              |                    |                |
|-----------------------|--------------|--------------------|----------------|
| Query Match           | 42.1%        | Score 539;         | DB 4;          |
| Best Local Similarity | 56.0%;       | Pred. No. 9.4e-54; | Length 212;    |
| Matches 103;          | Conservative | 30;                | Mismatches 51; |
|                       |              |                    | Indels 0;      |
|                       |              |                    | Gaps 0;        |

|    |  |     |                                                                       |     |
|----|--|-----|-----------------------------------------------------------------------|-----|
| QY |  | 42  | FLPKFVLVGSAGTGKSCLLJHQFIENFKODSNHTIGVEGSRVVNVGGTKVLQIWDTAG            | 101 |
|    |  |     | : : : : :           : : : : :           : : : : :                     |     |
| Db |  | 5   | YLKYIIIGDTGVGKSCLLLQFTDKRFQPVDHLTIGVEGARMITDGOIKLIQIWDTAG             | 64  |
|    |  |     | : : : : :           : : : : :           : : : : :                     |     |
| QY |  | 102 | QEFRSVRTSYRGAGALLIVDYITREYYNSLAAMLTDARTLASPNVILCGNKKLD                | 161 |
|    |  |     | : : : : :                     : : : : :           : : : : :           |     |
| Db |  | 65  | QESFRSITSYYRGAGALLIVDYITRDFTFNHLTTLWEDAROHNSNMVIMLGNSDLE              | 124 |
|    |  |     | : : : : :                     : : : : :           : : : : :           |     |
| QY |  | 162 | PEREVFFLRASRPAQENMLFLETSAITGENVBEEAFKCAITILNKIDSGELDPERMGS            | 221 |
|    |  |     | : : : : :           : : : : :           : : : : :                     |     |
| Db |  | 125 | SREVKKESGEAPAREGHGLIFMETSAKTASNVEEAEINTAKEYEKIQGVFDINNEANG            | 184 |
|    |  |     | : : : : :           : : : : :           : : : : :           : : : : : |     |
| QY |  | 222 | IQYG 225                                                              |     |
|    |  |     | : : : : :                                                             |     |
| Db |  | 185 | IKIG 188                                                              |     |

## RESULT 4

US-08-531-525-18  
Sequence 18, Application US/08531525  
Patent No. 5840683  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 5840683le, John F.  
APPLICANT: Abajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
TITLE OF INVENTION: of P21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/531,525  
FILING DATE: 21-SEP-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 37-94  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 212 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus norvegicus  
US-08-531-525-18

Query Match 41.9%; Score 536; DB 2; Length 212;  
Best Local Similarity 55.4%; Pred. No. 2.1e-53;  
Matches 102; Conservative 31; Mismatches 51; Indels 0; Gaps 0;  
QY 42 FLFKELVTSAGTGSCLLHQFIENKFKQDSNHTTGVFGSRVNVGGTKVQLQIWDTAG 101  
DB 5 YLFKVIIGDTGKSGCLLQFTDKRFQPVHDLTMGVEFGARMTIDGKQIKLQIWDTAG 64  
QY 102 QERFSVTSYRGAAGALLVVDITSRETYNSLAALWTDARTLASPNVIVILCGNKKOLD 161  
DB 65 QSEFSITNSYRGAAGALLVVDITRDTFNLTTWLEDAROHNSNVMVILGNKSDLE 124  
QY 162 PEREYTFLEASFAENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
DB 125 SRREYKKEGEFAREHGLIFMETSAKTASNVEEAFINTAKEIYEKIQSGVEDINNEANG 184  
QY 222 IOYG 225  
DB 185 IKIG 188

## RESULT 5

US-08-718-270A-18

Sequence 18, Application US/08718270A  
Patent No. 5910478  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 5910478le, John F.  
APPLICANT: Abajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptidomimetics Inhibiting  
TITLE OF INVENTION: the Oncogenic Action of P21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/718,270A  
FILING DATE: 20-SEP-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/531,525  
FILING DATE: 21-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/004,091  
FILING DATE: 21-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 78-95  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 212 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus norvegicus  
US-08-718-270A-18

Query Match 41.9%; Score 536; DB 2; Length 212;  
Best Local Similarity 55.4%; Pred. No. 2.1e-53;  
Matches 102; Conservative 31; Mismatches 51; Indels 0; Gaps 0;  
QY 42 FLFKELVTSAGTGSCLLHQFIENKFKQDSNHTTGVFGSRVNVGGTKVQLQIWDTAG 101  
DB 5 YLFKVIIGDTGKSGCLLQFTDKRFQPVHDLTMGVEFGARMTIDGKQIKLQIWDTAG 64  
QY 102 QERFSVTSYRGAAGALLVVDITSRETYNSLAALWTDARTLASPNVIVILCGNKKOLD 161  
DB 65 QSEFSITNSYRGAAGALLVVDITRDTFNLTTWLEDAROHNSNVMVILGNKSDLE 124  
QY 162 PEREYTFLEASFAENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
DB 125 SRREYKKEGEFAREHGLIFMETSAKTASNVEEAFINTAKEIYEKIQSGVEDINNEANG 184  
QY 222 IOYG 225  
DB 185 IKIG 188

RESULT 7  
US-08-718-270A-16  
; Sequence 16, Application US/08718270A  
; Patent No. 5910478

```

Query Match 39.7%; Score 508; DB 2; Length 210;
Best Local Similarity 54.9%; Pred. No. 3.3e-50;
Matches 101; Conservative 30; Mismatches 51; Indels 2; Gaps 2;

QY 42 FLKFLVIGSAGTCKSCLLHQFTENKFQDSNHTIGVEFSRVYNVGKTVKLQLWDTAG 101
 :||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5 YLFKYIIIGDTGVGKSCLLQLQFTDKRFQPVDLTIGVEFGARMITDGKIKLQIWDTAG 64

QY 102 QERFRSVTRYRGAACALLVYDITSRETYNSLAWLTDARTLASPNITWILCGNKKDLD 161
 ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 QESFRISIRSYRGAACALLVYDITRRDTFNHFLTWLEDAFQHSNNVMIVLIGNKSOLE 124

QY 162 PEREVTFLEASRAQENELMFLETSAITGENVEEAFKLCARTILNKIDSGELDPERMGSG 221
 | | | | | : | : | : | | | | | | | | | | | | | | | | | | | |
Db 125 -ARRVKKEGEAF-REHLIFMETSATAANVEEAFINTAKEIYQKIQDVFDINNEANG 182

QY 222 IOYG 225
 | . |
Db 183 IKIG 186

```



Sequence 17, Application US/08531525  
 Patent No. 5840683  
 GENERAL INFORMATION:  
 APPLICANT: Hlavka, Joseph J.  
 APPLICANT: Pincus, Matthew R.  
 APPLICANT: No. 5840683le, John F.  
 APPLICANT: Abajian, Henry B.  
 APPLICANT: Kende, Andrew S.  
 TITLE OF INVENTION: Peptidomimetics Inhibiting the Oncogenic Action  
 TITLE OF INVENTION: of P21 Ras  
 NUMBER OF SEQUENCES: 52  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Greenlee and Winner, P.C.  
 STREET: 5370 Mannheim Circle, Suite 201  
 CITY: Boulder  
 STATE: Colorado  
 COUNTRY: US  
 ZIP: 80303  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/531,525  
 FILING DATE: 21-SEP-1995  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ferber, Donna M.  
 REGISTRATION NUMBER: 33,878  
 REFERENCE/DOCKET NUMBER: 37-94  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (303) 499-8080  
 TELEFAX: (303) 499-8089  
 INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 208 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Oryctolagus cuniculus  
 US-08-531-525-17

Query Match 38.5%; Score 492.5; DB 2; Length 208;  
 Best Local Similarity 54.3%; Pred. No. 2e-48;  
 Matches 100; Conservative 30; Mismatches 51; Indels 3; Gaps 3;  
 QY 42 FLKFLVIGSAGTGKSCILLHOFIENKFKQDSNHTIGVEFGSRVYVGVGKTVKLIQWDTAG 101  
 DB 5 YLFYIIIGDTGVGKSCILLQFTDKRQPVHDLTIGVEFGARMITIDGKQIKLIQWDTA- 63  
 QY 102 QERFSVTRSYRGAAGALLVYDITSRETYNSLAWLTDARTLASPNVIVLCGNKKDLD 161  
 DB 64 QESRSI-RSYRG-AGALLVYDITRDTFNHLTWTLEDARQHSNSNMVIMLGNKSDLE 121  
 QY 162 PERVTFLASRFAQENELMLETSLTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
 DB 122 SRREVKEEGEAFAREHGLIFMETSAKTASNVEEAFINTAKEIYEKIOEGVFDINNEANG 181  
 QY 222 IQYG 225  
 DB 182 IKIG 185

RESULT 9  
 US-08-718-270A-17  
 Sequence 17, Application US/08718270A  
 Patent No. 5910478  
 GENERAL INFORMATION:  
 APPLICANT: Hlavka, Joseph J.

APPLICANT: Pincus, Matthew R.  
 APPLICANT: No. 5910478le, John F.  
 APPLICANT: Abajian, Henry B.  
 APPLICANT: Kende, Andrew S.  
 TITLE OF INVENTION: Peptidomimetics Inhibiting  
 TITLE OF INVENTION: the Oncogenic Action of P21 Ras  
 NUMBER OF SEQUENCES: 52  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
 STREET: 5370 Mannheim Circle, Suite 201  
 CITY: Boulder  
 STATE: Colorado  
 COUNTRY: US  
 ZIP: 80303  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/718,270A  
 FILING DATE: 20-SEP-1996  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/531,525  
 FILING DATE: 21-SEP-1995  
 PRIOR APPLICATION NUMBER: US 60/004,091  
 FILING DATE: 21-SEP-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ferber, Donna M.  
 REGISTRATION NUMBER: 33,878  
 REFERENCE/DOCKET NUMBER: 78-95  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (303) 499-8080  
 TELEFAX: (303) 499-8089  
 INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 208 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Oryctolagus cuniculus  
 US-08-718-270A-17

Query Match 38.5%; Score 492.5; DB 2; Length 208;  
 Best Local Similarity 54.3%; Pred. No. 2e-48;  
 Matches 100; Conservative 30; Mismatches 51; Indels 3; Gaps 3;  
 QY 42 FLKFLVIGSAGTGKSCILLHOFIENKFKQDSNHTIGVEFGSRVYVGVGKTVKLIQWDTAG 101  
 DB 5 YLFYIIIGDTGVGKSCILLQFTDKRQPVHDLTIGVEFGARMITIDGKQIKLIQWDTA- 63  
 QY 102 QERFSVTRSYRGAAGALLVYDITSRETYNSLAWLTDARTLASPNVIVLCGNKKDLD 161  
 DB 64 QESRSI-RSYRG-AGALLVYDITRDTFNHLTWTLEDARQHSNSNMVIMLGNKSDLE 121  
 QY 162 PERVTFLASRFAQENELMLETSLTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
 DB 122 SRREVKEEGEAFAREHGLIFMETSAKTASNVEEAFINTAKEIYEKIOEGVFDINNEANG 181  
 QY 222 IQYG 225  
 DB 182 IKIG 185

RESULT 10  
 US-08-531-525-15  
 Sequence 15, Application US/08531525  
 Patent No. 5840683

GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 5840683le, John F.  
APPLICANT: Abajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/531,525  
FILING DATE: 21-SEP-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 37-94  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 208 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-531-525-15  
Query Match 38.3%; Score 490; DB 2; Length 208;  
Best Local Similarity 53.8%; Pred. No. 3.8e-48;  
Matches 99; Conservative 31; Mismatches 50; Indels 4; Gaps 3;  
QY 42 FLKFLVIGSAGTCKSCLLHGFENKFKQDSNHTIGVEFGSRVNVGKTKVQLQIWDTAG 101  
Db 5 YLFYIIIGDTGVGKSCILLQFTDKRFQPVHDLTIGVEFGARMTIDGKQIKLQIWDTAG 64  
QY 102 QERFSRVSRYRGAAGALLVYDITSRETYNSLAWLTDARTLASPNVIVLGCNKKDLD 161  
Db 65 QESFRTSRYRGAAGALLVYDITRDTFNHLTWLEADARQHSNNVIMLGNKSDLE 124  
QY 162 PEREVTFLASRFAQENELMLETSLTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
Db 125 -RREVKEEGEAGA-EHLIFMETAKTA--SVEEAFINTAKEIYKIQGVGDINNEANG 180  
QY 222 IOYG 225  
Db 181 IKIG 184  
RESULT 11  
US-08-718-270A-15  
Sequence 15, Application US/08718270A  
Patent No. 5910478  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 5910478le, John F.

APPLICANT: Abajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptidomimetics Inhibiting  
the Oncogenic Action of P21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/718,270A  
FILING DATE: 20-SEP-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/531,525  
FILING DATE: 21-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/004,091  
FILING DATE: 21-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 78-95  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 208 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-718-270A-15  
Query Match 38.3%; Score 490; DB 2; Length 208;  
Best Local Similarity 53.8%; Pred. No. 3.8e-48;  
Matches 99; Conservative 31; Mismatches 50; Indels 4; Gaps 3;  
QY 42 FLKFLVIGSAGTCKSCLLHGFENKFKQDSNHTIGVEFGSRVNVGKTKVQLQIWDTAG 101  
Db 5 YLFYIIIGDTGVGKSCILLQFTDKRFQPVHDLTIGVEFGARMTIDGKQIKLQIWDTAG 64  
QY 102 QERFSRVSRYRGAAGALLVYDITSRETYNSLAWLTDARTLASPNVIVLGCNKKDLD 161  
Db 65 QESFRTSRYRGAAGALLVYDITRDTFNHLTWLEADARQHSNNVIMLGNKSDLE 124  
QY 162 PEREVTFLASRFAQENELMLETSLTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
Db 125 -RREVKEEGEAGA-EHLIFMETAKTA--SVEEAFINTAKEIYKIQGVGDINNEANG 180  
QY 222 IOYG 225  
Db 181 IKIG 184  
RESULT 12  
US-08-531-525-52  
Sequence 52, Application US/08531525  
Patent No. 5840683  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.

RESULT 14  
US-08-773-423-8  
; Sequence 8, Application US/08773423  
; Patent No. 5869291  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Bandman, Olga  
; TITLE OF INVENTION: NOVEL RAB PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA

RESULT 13  
US-08-718-270A-52  
; Sequence 52: Application US/08718270A  
; Patent No. 5910478  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5910478ie, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptidomimetics Inhibiting  
; TITLE OF INVENTION: the Oncogenic Action of p21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 14:58:57 ; Search time 5.93644 Seconds  
(without alignments)  
1761.695 Million cell updates/sec

Title: US-09-763-902B-8  
Perfect score: 1280  
Sequence: 1 MSVSLPLTVVVRDWTGIIH.....LRQLRQPSAQAVAPQPCGC 248

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues. 188354

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA.\*

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pap.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pap.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pap.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pap.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pap.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pap.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pap.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pap.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pap.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pap.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pap.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 1147  | 89.6        | 222    | 9     | US-09-764-868-1112 |
| 2          | 1147  | 89.6        | 225    | 9     | US-09-764-868-692  |
| 3          | 645.5 | 50.4        | 307    | 9     | US-09-764-868-1100 |
| 4          | 645.5 | 50.4        | 312    | 10    | US-09-925-302-783  |
| 5          | 539   | 42.1        | 212    | 10    | US-09-350-874-67   |
| 6          | 425.5 | 33.2        | 213    | 10    | US-09-988-974-8    |
| 7          | 424   | 33.1        | 222    | 10    | US-09-820-003A-4   |
| 8          | 421.5 | 32.9        | 217    | 10    | US-09-988-974-3    |
| 9          | 421.5 | 32.9        | 239    | 10    | US-09-925-301-1077 |
| 10         | 417.5 | 32.6        | 201    | 10    | US-09-967-736-3    |
| 11         | 417   | 32.6        | 201    | 10    | US-09-967-736-8    |
| 12         | 414.5 | 32.4        | 207    | 10    | US-09-794-257-8    |
| 13         | 413.5 | 32.3        | 213    | 10    | US-09-794-257-5    |
| 14         | 403.5 | 31.5        | 246    | 10    | US-09-925-302-534  |
| 15         | 400   | 31.2        | 190    | 10    | US-09-822-860-5    |
| 16         | 391.5 | 30.6        | 198    | 10    | US-09-794-257-16   |
| 17         | 391.5 | 30.6        | 198    | 10    | US-09-945-173-5    |
| 18         | 391.5 | 30.6        | 198    | 10    | US-09-972-529-4    |
| 19         | 389   | 30.4        | 218    | 10    | US-09-925-300-1571 |

20 379.5 29.6 226 9 US-09-764-868-684 Sequence 684, App  
21 378.5 29.6 222 9 US-09-764-868-1106 Sequence 1106, Ap  
22 377.5 29.5 223 10 US-09-817-199A-2 Sequence 2, Appli  
23 376 29.4 223 10 US-09-817-199A-4 Sequence 4, Appli  
24 374.5 29.3 212 10 US-09-817-198A-2 Sequence 2, Appli  
25 374.5 29.3 401 9 US-09-764-868-701 Sequence 701, App  
26 371.5 29.0 191 10 US-09-794-257-14 Sequence 14, Appl  
27 371.5 29.0 191 12 US-10-051-986-3 Sequence 3, Appli  
28 371.5 29.0 212 10 US-09-817-198A-4 Sequence 4, Appli  
29 371 29.0 162 10 US-09-834-765-766 Sequence 766, App  
30 363 28.4 133 9 US-09-764-868-1101 Sequence 1101, Ap  
31 361.5 28.2 218 10 US-09-817-198A-5 Sequence 5, Appli  
32 352 27.5 188 9 US-09-764-868-1120 Sequence 1120, Ap  
33 343.5 26.8 208 9 US-10-108-608-45 Sequence 45, Appl  
34 336 26.2 217 10 US-09-925-300-1364 Sequence 1364, Ap  
35 334.5 26.1 255 9 US-09-764-868-686 Sequence 686, App  
36 314 24.5 216 10 US-09-945-173-10 Sequence 10, Appl  
37 310.5 24.3 201 10 US-09-822-860-2 Sequence 2, Appli  
38 307 24.0 221 9 US-09-764-868-702 Sequence 702, App  
39 307 24.0 222 9 US-09-764-868-1121 Sequence 1121, Ap  
40 303 23.7 208 10 US-09-925-302-629 Sequence 629, App  
41 300 23.4 832 10 US-09-834-765-2 Sequence 2, Appli  
42 299.5 23.4 201 10 US-09-988-974-9 Sequence 9, Appli  
43 296.5 23.2 201 10 US-09-988-974-5 Sequence 5, Appli  
44 296.5 23.2 209 10 US-09-864-761-42996 Sequence 42996, A  
45 293 22.9 157 10 US-09-834-765-764 Sequence 764, App

#### ALIGNMENTS

RESULT 1  
US-09-764-868-1112  
; Sequence 1112, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PZ32  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1112  
; LENGTH: 222  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-868-1112

Query Match 89.6%; Score 1147; DB 9; Length 222;  
Best Local Similarity 100.0%; Pred. No. 1.4e-108; Indels 0; Gaps 0;  
Matches 222; Conservative 0; Mismatches 0;

QY 27 SLPVGIPDFGSIWSDFLKFLVIGSAGTGKSCLLHQFIENKFKQDSNHTIGVEFGSRVYN 86  
Db 1 SLPVGIPDFGSIWSDFLKFLVIGSAGTGKSCLLHQFIENKFKQDSNHTIGVEFGSRVYN 60  
QY 87 VGGKTVKLIQINDTAGQERFSVTSYIRGAAGALLVYDITSRETYNSLAAILTDARTLAS 146  
Db 61 VGGKTVKLIQINDTAGQERFSVTSYIRGAAGALLVYDITSRETYNSLAAILTDARTLAS 120  
QY 147 PNIVILCGNKKDLDPERVTTFLEASRFAQENELMFLFETSALTGENVEEAFKLCARTILN 206  
Db 121 PNIVILCGNKKDLDPERVTTFLEASRFAQENELMFLFETSALTGENVEEAFKLCARTILN 180  
QY 207 KIDSGELDPERMGSIGYGDASLRQLRQPSAQAVAPQPCGC 248  
Db 181 KIDSGELDPERMGSIGYGDASLRQLRQPSAQAVAPQPCGC 222

RESULT 2  
US-09-764-868-692

; Sequence 692, Application US/09764868  
; Patent No. US20020168711A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT232

; CURRENT APPLICATION NUMBER: US/09/764,868

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1510

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 692

; LENGTH: 225

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-868-692

Query Match 89.6%; Score 1147; DB 9; Length 225;

Best Local Similarity 100.0%; Pred. No. 1.4e-108;

Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 SLPGVDPFGSTWSDFLFKFLVIGSAGTGKSCLLHQFIEKFKQDSNHTIGVEFGSRVYN 86

Db 4 SLPGVDPFGSIWDFLFKFLVIGSAGTGKSCLLHQFIEKFKQDSNHTIGVEFGSRVYN 63

QY 87 VGGKTVKLIQIWDTAGQERFSVTSYRGAAGALLVYDITSRETNLSLAWLTDARTLAS 146

Db 64 VGGKTVKLIQIWDTAGQERFSVTSYRGAAGALLVYDITSRETNLSLAWLTDARTLAS 123

QY 147 PRNVILCGNKKDLDPEREVTFLEASRAQENELMFLTSALTGENVEEAFKLCARTILN 206

Db 124 PRNVILCGNKKDLDPEREVTFLEASRAQENELMFLTSALTGENVEEAFKLCARTILN 183

QY 207 KIDSGELDPERMGSIGIQGDASLRQLRPSAQAVAPQCGC 248

Db 184 KIDSGELDPERMGSIGIQGDASLRQLRPSAQAVAPQCGC 225

RESULT 3

US-09-764-868-1100

; Sequence 1100, Application US/09764868

; Patent No. US20020168711A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT232

; CURRENT APPLICATION NUMBER: US/09/764,868

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1510

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1100

; LENGTH: 307

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-868-1100

Query Match 50.4%; Score 645.5; DB 9; Length 307;

Best Local Similarity 58.0%; Pred. No. 1.4e-57;

Matches 123; Conservative 33; Mismatches 45; Indels 11; Gaps 3;

QY 42 FLFKFLVIGSAGTGKSCLLHQFIEKFKQDSNHTIGVEFGSRVNVGKTVKLIQIWDTAG 101

Db 102 YIFKYIIIGDMGVKSCLLHQFTEKKFADCPHTIGVEGTRIEVSGOKIKLIQIWDTAG 161

QY 102 QERFSVTSYRGAAGALLVYDITSRETNLSLAWLTDARTLASPNIVILCGNKKDL 161

Db 162 QERFSVTSYRGAAGALLVYDITSRETNLSLAWLTDARTLASPNIVILCGNKKDL 221

QY 162 PEREVTLEASRAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221

Db 222 AQORDVTEEAQFAENGCLLFLEASAKTGENVEDAFLEAAKKIYQNIQDGSIDLNAESG 281

QY 222 IOYGDASLRQLRPSAQAVA-PQP-----CGC 248

Db 282 VQHKPSA-----POGRLTSEPPQREGCGC 307

RESULT 4

US-09-925-302-783

; Sequence 783, Application US/09925302

; Patent No. US20020044941A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA104

; CURRENT APPLICATION NUMBER: US/09/925,302

; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05918

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 896

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 783

; LENGTH: 312

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-925-302-783

Query Match 50.4%; Score 645.5; DB 10; Length 312;

Best Local Similarity 58.0%; Pred. No. 1.5e-57;

Matches 123; Conservative 33; Mismatches 45; Indels 11; Gaps 3;

QY 42 FLFKFLVIGSAGTGKSCLLHQFIEKFKQDSNHTIGVEFGSRVNVGKTVKLIQIWDTAG 101

Db 107 YIFKYIIIGDMGVKSCLLHQFTEKKFADCPHTIGVEGTRIEVSGOKIKLIQIWDTAG 166

QY 102 QERFSVTSYRGAAGALLVYDITSRETNLSLAWLTDARTLASPNIVILCGNKKDL 161

Db 167 QERFSVTSYRGAAGALLVYDITSRETNLSLAWLTDARTLASPNIVILCGNKKDL 226

QY 162 PEREVTLEASRAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221

Db 227 AQORDVTEEAQFAENGCLLFLEASAKTGENVEDAFLEAAKKIYQNIQDGSIDLNAESG 286

QY 222 IOYGDASLRQLRPSAQAVA-PQP-----CGC 248

Db 287 VQHKPSA-----POGRLTSEPPQREGCGC 312

RESULT 5

US-09-350-874-67

; Sequence 67, Application US/09350874

; Patent No. US20020019020A1

; GENERAL INFORMATION:

; APPLICANT: Rhodes, Kenneth

; TITLE OF INVENTION: METHODS FOR TREATING CARDIOVASCULAR DISORDERS

; FILE REFERENCE: MNI-069

; CURRENT APPLICATION NUMBER: US/09/350,874

; CURRENT FILING DATE: 1999-07-09

; EARLIER APPLICATION NUMBER: USSN 60/110,277

; EARLIER FILING DATE: 1998-11-30

; EARLIER APPLICATION NUMBER: USSN 60/110,033

; EARLIER FILING DATE: 1998-11-25

; EARLIER APPLICATION NUMBER: USSN 60/109,333

; EARLIER FILING DATE: 1998-11-20

; EARLIER APPLICATION NUMBER: USSN 09/298,731

; EARLIER FILING DATE: 1999-04-23

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 67

; LENGTH: 212

; TYPE: PRT

; ORGANISM: Rattus sp.

US-09-350-874-67

Query Match 42.1%; Score 539; DB 10; Length 212;  
Best Local Similarity 56.0%; Pred. No. 5.3e-47;  
Matches 103; Conservative 30; Mismatches 51; Indels 0; Gaps 0;  
QY 42 FLFKFLVIGSAGTGSCLLHQFIENKFKQDSNHTIGVEGSRVNVGGKTVKLIQIWDTAG 101  
DB 5 YLFKYIIIGDTGVKSCLLQFFDKRFQPHVDTIGVEGARMITIDGKIQLQIWDTAG 64  
QY 102 QERFRSVTRSYRGAAGALLVYDITSRETYNSLAAWLTARTLASPNIVVILCGNKKDL 161  
DB 65 QESFRSITRSYRGAAGALLVYDITSRETYNSLAAWLTARTLASPNIVVILCGNKKDL 124  
QY 162 PEREVTFLASRAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 221  
DB 125 SRREVKEEGEAPAREHGLIFMETSAKTASNVEEAFINTAKEIYEKIQGVFDINNEANG 184  
QY 222 IQYG 225  
DB 185 IKIG 188

RESULT 6

US-09-988-974-8  
Sequence 8, Application US/09988974  
Patent No. US20020090712A1  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
Bandman, Olga

TITLE OF INVENTION: NOVEL RAB PROTEINS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/988.974  
FILING DATE: 19-NOV-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/215,887  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PP-0183 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 213 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436001  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-988-974-8

Query Match

33.2%; Score 425.5; DB 10; Length 213;

Best Local Similarity

Matches 93; Conservative 28; Mismatches 78; Indels 1; Gaps 1;

QY 41 DFLFKFLVIGSAGTGSCLLHQFIENKFKQDSNHTIGVEGSRVNVGGKTVKLIQIWDTA 100  
DB 10 NFVFKVVLIGESGVGTNLLSRTRENEFSHDSRTTIGVEFSTVLLGTAARVAKIWDTA 69  
QY 101 GQERFRSVTRSYRGAAGALLVYDITSRETYNSLAAWLTARTLASPNIVVILCGNKKDL 160  
DB 70 GLERYRAITSYRGAAGALLVYDITSRETYNSLAAWLTARTLASPNIVVILCGNKKDL 129  
QY 161 DPREVTFLEASRAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 220  
DB 130 SQAREVPTTEARFAENGLIFLTSALDSTNVELAFETVTKEIFAKV-SKQIQNSPRSN 188  
QY 221 GIOYGDASLRQLRQPSAQA 240  
DB 189 AIALGSAQAGQEPGQKRA 208

RESULT 7

US-09-820-003A-4  
Sequence 4, Application US/09820003A  
Patent No. US20020142382A1  
GENERAL INFORMATION:  
APPLICANT: MERKULOV, Gennady et al.

TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
PROTEINS, AND USES THEREOF  
FILE REFERENCE: CLO01196  
CURRENT APPLICATION NUMBER: US/09/820,003A  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 4

LENGTH: 222  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-820-003A-4

Query Match 33.1%; Score 424; DB 10; Length 222;

Best Local Similarity 45.5%; Pred. No. 2.5e-35;  
Matches 85; Conservative 33; Mismatches 59; Indels 10; Gaps 1;

QY 41 DFLFKFLVIGSAGTGSCLLHQFIENKFKQDSNHTIGVEGSRVNVGGKTVKLIQIWDTA 100  
DB 26 DYLFKLLIGDSGVKSCLLRFADDTYTESYISTGVDFKIRTIELDKTKIKLIQIWDTA 85  
QY 101 GQERFRSVTRSYRGAAGALLVYDITSRETYNSLAAWLTARTLASPNIVVILCGNKKDL 160  
DB 86 GQERFRITTSYRGAAGIIVYDITDQESFNKVKQWLOEIDRYASENVNKLIVGNKCDL 145  
QY 161 DPREVTFLEASRAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMGS 220  
DB 146 TTKKVVDYTTAKEFADSLGIPFLTSAKNATNVEQSEWNTAAEI-----KKRMGP 195  
QY 221 GIOYGDA 227  
DB 196 GATAGGA 202

RESULT 8

US-09-988-974-3  
Sequence 3, Application US/09988974  
Patent No. US20020090712A1  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
Bandman, Olga

TITLE OF INVENTION: NOVEL RAB PROTEINS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.

```
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/988,974
FILING DATE: 19-NO. US20020090712A1-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/215,887
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J. 749
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0183 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Consensus
CLONE: Consensus
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-988-974-3
Query Match 32.9%; Score 421.5; DB 10; Length 217;
Best Local Similarity 46.5%; Pred. No. 4.4e-35;
Matches 93; Conservative 27; Mismatches 79; Indels 1; Gaps 1;
QY 41 DFLFKVLVSGAGTGSCLLHQFIENKFKQDSNHTTIGVEFGSRVNVGKTVKQLQIWDTA 100
DB 14 NFVKVVLIGESGVGKTNLLSRFNEFSHDSRTTIGVEFTVMTGTAARQAQIWDTA 73
QY 101 GOERFRSVTSYRGAAGALLVDITSRITYNSLAAMLTDARTLASPNIVVILCGNKKDL 160
DB 74 GLERYRAITSAYRGAAGALLVFDLTKHQTAVVERWLKELYDHAETIVVVLVGNKSDL 133
QY 161 DPERVTFLEASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMS 220
DB 134 SOGREVPTTEARFAENGLLFLTSALDSTNVELAFETVLKEIFAKVSKQRQNSIRT-N 192
QY 221 GIOYGDASLRQLRQPSAQA 240
DB 193 AITLGSQAQXGQEPGPKRA 212
RESULT 9
US-09-925-301-1077
Sequence 1077, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIORITY APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/988,974
FILING DATE: 19-NO. US20020090712A1-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/215,887
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J. 749
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0183 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Consensus
CLONE: Consensus
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-988-974-3
Query Match 32.9%; Score 421.5; DB 10; Length 217;
Best Local Similarity 46.5%; Pred. No. 4.4e-35;
Matches 93; Conservative 27; Mismatches 79; Indels 1; Gaps 1;
QY 41 DFLFKVLVSGAGTGSCLLHQFIENKFKQDSNHTTIGVEFGSRVNVGKTVKQLQIWDTA 100
DB 14 NFVKVVLIGESGVGKTNLLSRFNEFSHDSRTTIGVEFTVMTGTAARQAQIWDTA 73
QY 101 GOERFRSVTSYRGAAGALLVDITSRITYNSLAAMLTDARTLASPNIVVILCGNKKDL 160
DB 74 GLERYRAITSAYRGAAGALLVFDLTKHQTAVVERWLKELYDHAETIVVVLVGNKSDL 133
QY 161 DPERVTFLEASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMS 220
DB 134 SOGREVPTTEARFAENGLLFLTSALDSTNVELAFETVLKEIFAKVSKQRQNSIRT-N 192
QY 221 GIOYGDASLRQLRQPSAQA 240
DB 193 AITLGSQAQXGQEPGPKRA 212
RESULT 9
US-09-925-301-1077
Sequence 1077, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: RAB PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/967,736
FILING DATE: 28-Sep-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/154,602
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS: single
```



```

;
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: L1V1TUT04
; CLONE: 2514506
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-967-736-3

Query Match 32.6%; Score 417.5; DB 10; Length 201;
Best Local Similarity 42.7%; Pred. No. 1e-34;
Matches 90; Conservative 32; Mismatches 70; Indels 19; Gaps 3;

QY 41 DFLFKFLVIGSAGTGSCLLHQFIENKFKQDSNHTIGVEFGSRVNVGGTKVTKLQIWDTA 100
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6 DYLFKLLIGDSGVGKSCLLRFADDYITESTYSTIGVDFKIRTIEDGKTIKLIQIWDTA 65

QY 101 GQFRFRSVTSYRGAAGALLVYDITSRETYNSLAAMLTDARTLASPNIVVILCGNKKDL 160
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 GQERFRVTSYRGAAGIIVYDVTDQESYANYKQMLQEIDRYASENVNKLVLGNKSDL 125

QY 161 DPREVTFLEASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMG 220
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 TTRKVVDTNTAKEFADSLGPFLETSKATNATNVEQAFMTMAAEI-----KRRMP 175

QY 221 GIQYGDASLRQLRQPR---SAQAVAPQCGC 248
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176 GAASGG-----ERNLKDITPVKPGAGGC 200.

RESULT 11
US-09-967-736-8
; Sequence 8, Application US/09967736
; Patent No. US20020103340A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; Corley, Neil C.
; Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/967,736
; FILING DATE: 28-Sep-2001
; APPLICATION DATA:
; APPLICATION NUMBER: 09/154,602
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0367 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 57006

```

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-967-736-8

Query Match 32.6%; Score 417; DB 10; Length 201;
Best Local Similarity 46.5%; Pred. No. 1.1e-34;
Matches 86; Conservative 29; Mismatches 60; Indels 10; Gaps 1;

QY 41 DFLFKFLVIGSAGTGSCLLHQFIENKFKQDSNHTIGVEFGSRVNVGGTKVTKLQIWDTA 100
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6 DYLFKLLIGDSGVGKSCLLRFADDYITESTYSTIGVDFKIRTIEDGKTIKLIQIWDTA 65

QY 101 GQFRFRSVTSYRGAAGALLVYDITSRETYNSLAAMLTDARTLASPNIVVILCGNKKDL 160
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 GQERFRVTSYRGAAGIIVYDVTDQESYANYKQMLQEIDRYASENVNKLVLGNKSDL 125

QY 161 DPREVTFLEASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMG 220
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 TTRKVVDTNTAKEFADSLGPFLETSKATNATNVEQAFMTMAAEI-----KRRMP 175

QY 221 GIQYGDASLRQLRQPRSA 238
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 186 G---GPVKITENRSKKT 200

RESULT 13
US-09-794-257-5
; Sequence 5, Application US/09794257
; Patent No. US20020009804A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32705, 27423, 32700, 32712, No. US20020009804A1el

Query Match 32.4%; Score 414.5; DB 10; Length 207;
Best Local Similarity 40.9%; Pred. No. 2.1e-34;
Matches 81; Conservative 46; Mismatches 68; Indels 3; Gaps 1;

QY 41 DFLFKFLVIGSAGTGSCLLHQFIENKFKQDSNHTIGVEFGSRVNVGGTKVTKLQIWDTA 100
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6 DYLFKLLIGDSGVGKTCLLFRSEDAFTTFTSTIGIDFKIRTIEDGKTIKLIQIWDTA 65

QY 101 GQFRFRSVTSYRGAAGALLVYDITSRETYNSLAAMLTDARTLASPNIVVILCGNKKDL 160
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 GQERFRVTSYRGAAGIIVYDITNEKSPDNIRNIRNIEEHASSDVERMILGNKCDM 125

QY 161 DPREVTFLEASRFAQENELMFLTSALTGENVEEAFKLCARTILNKIDSGELDPERMG 220
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 NDKRQVSKERGEKLAIDYGIKFELETSKASSANVEEAFFTLARDIMTKLRKNMDSNSAGA 185

QY 221 GIQYGDASLRQLRQPRSA 238
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 186 G---GPVKITENRSKKT 200

RESULT 13
US-09-794-257-5
; Sequence 5, Application US/09794257
; Patent No. US20020009804A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32705, 27423, 32700, 32712, No. US20020009804A1el

```

;; TITLE OF INVENTION: Human G-Proteins  
;; FILE REFERENCE: 35800/209285  
;; CURRENT APPLICATION NUMBER: US/09/794,257  
;; CURRENT FILING DATE: 2001-02-27  
;; PRIOR APPLICATION NUMBER: 60/185,606  
;; PRIOR FILING DATE: 2000-02-29  
;; NUMBER OF SEQ ID NOS: 16  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 5  
;; LENGTH: 213  
;; TYPE: PRT  
;; ORGANISM: homo sapiens  
US-09-794-257-5

Query Match 32.3%; Score 413.5; DB 10; Length 213;  
Best Local Similarity 44.0%; Pred. No. 2.8e-34;  
Matches 84; Conservative 40; Mismatches 64; Indels 3; Gaps 3;  
QY 37 SIWSDFLKFLVIGSAGTGKSCLLHQFIENKFKQDSNHTIGVEFGSRVNV-GGKTVKQLQ 95  
DB 3 AIWL-YQFLIVIGDSTVGSCLIRRTGEGFAQVSDPTVGVDFFSLVLEPGKRIKLO 61  
QY 96 IWDTAGQERFSVTRSYRGAAGALLVYDITSRETYNSLAAMLTDARTLASP-NIVVILC 154  
DB 62 IWDTAGQERFSVTRSYRGSVGLLLFDITNRRSFQNVHLEETKVHVQYQIVFVLV 121  
QY 155 GNKKDLPDPEVTFLEASRFAQENELMFLETSALTGENVEAFKLCARTILNKIDSGELD 214  
DB 122 GHKCDLDTQVTRHEAEKLAAYGMKYIETSDAINVEKAFETDLTRDIYELVKRGEIT 181  
QY 215 PERMSGIOYK 225  
DB 182 IQEGWEGKSG 192

RESULT 14  
US-09-925-302-534  
;; Sequence 534, Application: US/09925302  
;; Patent No. US20020044941A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
;; FILE REFERENCE: PA104  
;; CURRENT APPLICATION NUMBER: US/09/925,302  
;; CURRENT FILING DATE: 2001-08-10  
;; PRIOR APPLICATION NUMBER: PCT/US00/05918  
;; PRIOR FILING DATE: 2000-03-08  
;; PRIOR APPLICATION NUMBER: 60/124,270  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 896  
;; SOFTWARE: PatentIn ver. 2.0  
;; SEQ ID NO 534  
;; LENGTH: 246  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-925-302-534

Query Match 31.5%; Score 403.5; DB 10; Length 246;  
Best Local Similarity 44.8%; Pred. No. 3.5e-33;  
Matches 81; Conservative 38; Mismatches 57; Indels 5; Gaps 2;  
QY 41 DFLKFLVIGSAGTGKSCLLHQFIENKFKQDSNHTIGVEFGSRVNVGGKTVKLOIWDTA 100  
DB 49 DHLKFLLLIGDSGVGKTCLLIRFAEDNENNTYISTIGIDFKIRTVDIEGKKIKLQVWDTA 108  
QY 101 GOERFSVTRSYRGAAGALLVYDITSRETYNSLAAMLTDARTLASPNIVVILCGNKKDL 160  
DB 109 GOERKTTITAYRGAMGIIIVYDITDEKSFENIONWWSIKENASAGVERILLGNKCDM 168  
QY 161 DPEREVTFLEASRFAQENELMFLETSALTGENVEAFKLCARTILNKIDSGELDPERMS 220  
DB 169 EAKRVRQEOADKLAREHGIRFFETSAKSSNMVDEAFSSLLARDILK--SG---GRRSGN 223

QY 221 G 221  
DB 224 G 224

RESULT 15  
US-09-822-860-5  
;; Sequence 5, Application US/09822860  
;; Patent No. US20020146795A1  
;; GENERAL INFORMATION:  
;; APPLICANT: ZHU, Shiaooping et al.  
;; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
;; TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
;; FILE REFERENCE: CL001214  
;; CURRENT APPLICATION NUMBER: US/09/822,860  
;; CURRENT FILING DATE: 2001-04-02  
;; NUMBER OF SEQ ID NOS: 5  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 5  
;; LENGTH: 190  
;; TYPE: PRT  
;; ORGANISM: Discopyge ommata  
US-09-822-860-5  
Query Match 31.2%; Score 400; DB 10; Length 190;  
Best Local Similarity 45.0%; Pred. No. 5.5e-33;  
Matches 76; Conservative 40; Mismatches 53; Indels 0; Gaps 0;  
QY 41 DFLKFLVIGSAGTGKSCLLHQFIENKFKQDSNHTIGVEFGSRVNVGGKTVKLOIWDTA 100  
DB 1 DYLKFLLLIGDSGVGKTCLLFRPSEDAFTTISTIGIDFKIRTVELDGKKIKLQIWDTA 60  
QY 101 GOERFSVTRSYRGAAGALLVYDITSRETYNSLAAMLTDARTLASPNIVVILCGNKKDL 160  
DB 61 GOERFRTITAYRGAMGIMKVVYDITNEKSPDNKNNIRNIEEHASSOVERMILGNKCDM 120  
QY 161 DPEREVTFLEASRFAQENELMFLETSALTGENVEAFKLCARTILNKID 209  
DB 121 NEKQVSKERGEKLAIDYGIKFLKLETSKSSINVEAFITLARDIMTKLN 169

Search completed: March 4, 2003, 15:07:29  
Job time : 6.93644 secs